

**Modellierung von Metalloenzymen:
3D-QSAR-Untersuchungen an Carboanhydrase-Isoenzymen
und
virtuelles Screening nach Peptiddeformylase-Inhibitoren**

Dissertation
zur
Erlangung des
Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem
Fachbereich Pharmazie
der PHILIPPS-UNIVERSITÄT MARBURG
vorgelegt von

Alexander Hillebrecht
aus Friedberg

Marburg/Lahn 2007

Vom Fachbereich Pharmazie der Philipps-Universität Marburg
als Dissertation angenommen am:

20.12.2007

Erstgutachter:

Prof. Dr. G. KLEBE

Zweitgutachter:

Prof. Dr. C. A. SOTRIFFER

Tag der mündlichen Prüfung:

21.12.2007

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn Prof. Dr. G. KLEBE am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von Juni 2004 bis Oktober 2007 durchgeführt.

All models are wrong, but some are useful.

GEORGE E. P. BOX

Abkürzungsverzeichnis

n D	n -dimensional
Abb.	Abbildung
Abk.	Abkürzung
AMPA	α -Amino-3-hydroxy-5-methyl-4-isoxazolpropionsäure
ANN	<i>Artificial Neural Network</i>
AUC	<i>Area Under the Curve</i>
CA	Carboanhydrase
CAI	Carboanhydraseinhibitor
CARP	<i>Carbonic Anhydrase Related Protein</i>
COX	Cyclooxygenase
CPCA	<i>Consensus Principal Component Analysis</i>
Da	Dalton
EKG	Elektrokardiogramm
Fig.	<i>Figure</i>
GABA	γ -Aminobuttersäure
HDAC	Histondeacetylase
5-HT	5-Hydroxytryptamin, Serotonin
hERG	<i>Human Ether-A-Go-Go Related Gene</i>
INN	<i>International Nonproprietary Name</i> , Internationaler Freiname
Kap.	Kapitel
MACCS	<i>Molecular Access System</i>
MDR	<i>Multiple Drug Resistance</i>
MLR	multiple lineare Regression
NSAR	Nicht-steroidale Antirheumatika
OLS	<i>Ordinary Least Squares</i> , Synonym für MLR
p.	<i>page</i>
PCA	<i>Principal Component Analysis</i> , Hauptkomponentenanalyse
PCR	<i>Principal Component Regression</i> , Hauptkomponentenregression
PDB	<i>Protein Data Bank</i>

PDF	(AtPDF, Peptiddeformylase (aus <i>Arabidopsis thaliana</i> , EcPDF, HsPDF, <i>Escherichia coli</i> , <i>Homo sapiens</i> , <i>Plasmodium</i> PfPDF)
P-gp	P-Glykoprotein
PLS	<i>Partial Least Squares</i> oder <i>Projection to Latent Structures</i>
QSAR	<i>Quantitative Structure Activity Relationships</i> , Quantitative Struktur-Wirkungs-Beziehungen
RMSD	<i>Root Mean-Square Deviation</i>
ROC	<i>Receiver Operating Characteristic</i>
s.	siehe
S.	Seite
SPL	<i>SYBYL Programming Language</i>
SVL	<i>Scientific Vector Language</i>
SVM	<i>Support Vector Machine</i>
Tab.	Tabelle
UAW	Unerwünschte Arzneimittelwirkung
VSA	van-der-Waals Surface Area
VS	Virtuelles Screening
WDI	<i>World Drug Index</i>
ZNS	Zentralnervensystem

Inhaltsverzeichnis

1	Einleitung und Problemstellung	1
1.1	Selektivität und Arzneistoffentwicklung	1
1.2	Virtuelles Screening zur Leitstrukturfindung	3
2	Selektivitätsvorhersage mittels chemometrischer Methoden	4
2.1	CoMFA und CoMSIA	4
2.2	AFMoC	8
2.3	3D-QSAR zur Selektivitätsvorhersage	13
2.4	Alignmentverfahren	14
2.4.1	Manuelle Platzierung und anschließende Minimierung in der Bin- detasche	15
2.4.2	Docking: FlexX, AutoDock, GOLD	16
2.4.3	FlexS	18
2.5	GRID/CPCA	19
2.6	Beispiele aus der Literatur	23
3	Carboanhydrase als Modellsystem	25
3.1	Klassifizierung, Struktur und katalytischer Mechanismus	25
3.1.1	α -Carboanhydrasen	26
3.1.2	β -Carboanhydrasen	30
3.1.3	γ -Carboanhydrasen	32
3.1.4	δ -Carboanhydrasen und cadmiumhaltige CAs	33
3.2	Isoenzyme der humanen α -CAs	33
3.2.1	CA I	34
3.2.2	CA II	34
3.2.3	CA III	35
3.2.4	CA IV	35
3.2.5	CA VA und VB	37
3.2.6	CA VI	38
3.2.7	CA VII	38
3.2.8	CA IX und XII	39
3.2.9	CA XIII, XIV und XV	40
3.2.10	CA-verwandte Proteine (CARP)	42

3.3	Carboanhydraseinhibitoren	42
3.3.1	Substanzklassen	43
3.3.2	Therapeutisch genutzte Substanzen mit CA-inhibitorischer Wirkung	47
4	3D-QSAR-Studien zur Identifizierung affinitäts- und selektivitätsbestimmender Faktoren	52
4.1	Einleitung	52
4.2	Target Proteins	52
4.3	Methods	54
4.3.1	Dataset and Alignment	54
4.3.2	CoMFA and CoMSIA Analyses	56
4.3.3	AFMoC Analysis	57
4.3.4	3D-QSAR Based Selectivity Maps	58
4.3.5	Protein Based Selectivity Maps	59
4.4	Results and Discussion	60
4.4.1	Statistical Results	60
4.4.2	Visualization of the Property Maps	62
4.4.3	Differential GRID Maps and CPCA	68
4.4.4	Synopsis of the Approaches	71
4.5	Conclusions	74
5	Untersuchung der Eignung von 3D-QSAR-Modellen für Datenbankscreening	80
5.1	Einleitende Bemerkung	80
5.2	Abstract	80
5.3	Introduction	80
5.4	Methods	83
5.4.1	Dataset and Preparation	84
5.4.2	3D-QSAR Analyses	85
5.4.3	1D- and 2D-QSAR Analyses	85
5.4.4	Evaluation of Predictivity	86
5.5	Results and Discussion	88
5.5.1	Comparison of 3D- and 2D-QSAR Models in Terms of Internal Predictivity	88
5.5.2	External Numerical Predictivity: Correlation Coefficient	89

5.5.3	External Numerical Predictivity: Experimental versus Predicted Plots	91
5.5.4	External Categorical Predictivity: Sensitivity, Specificity, Hit Rates	97
5.5.5	External Categorical Predictivity: ROC Plots	100
5.5.6	External Categorical Predictivity: Classical Enrichment Plots .	101
5.5.7	Interpretation of the MACCS Models	104
5.6	Conclusion	107
6	Bewertung virtueller Ligandbibliotheken durch QSAR-Modelle	110
6.1	Einleitung	110
6.2	Zusammensetzung der Ligandbibliothek	111
6.3	Datenbankerzeugung	111
6.3.1	para-Substituierte Phenylsulfonamide – Grundgerüst B	112
6.3.2	Thiadiazolsulfonamide – Grundgerüst A	112
6.3.3	Benzothiazolsulfonamide – Grundgerüst C	112
6.3.4	Grundgerüste H und I	113
6.4	Bewertung der virtuellen Bibliothek mit QSAR-Modellen	113
6.5	Ergebnisse	114
6.6	Ausblick	118
7	Virtuelles Screening nach neuartigen Inhibitoren der Peptiddeformylasen (PDF)	119
7.1	Struktur, Klassifizierung und katalytischer Mechanismus der PDF . . .	119
7.2	Bekannte Inhibitoren der PDF aus der Literatur	123
7.3	CATALYST – theoretischer Hintergrund	125
7.4	Generierung von Pharmakophorhypothesen für PDF-Inhibitoren	127
7.5	Selektivitätsmodellierung gegenüber humaner PDF	130
7.6	Validierung und Datenbanksuche	131
7.7	Postprozessierung und Auswahl der Screening-Hits	136
7.8	Identifizierung weiterer potentieller Hydroxamatliganden	145
7.9	Ergebnisse und Diskussion	147
8	Anhang	155
8.1	Verwendete Hardware	155
8.2	Verwendete Software	155
8.3	Statistische Parameter der PLS-Analyse	156

9 Zusammenfassung, Summary	159
9.1 Zusammenfassung	159
9.2 Summary	161
Literaturverzeichnis	163

1 Einleitung und Problemstellung

1.1 Selektivität und Arzneistoffentwicklung

In den frühen Phasen der Entwicklung neuer Arzneistoffe wird oft zwischen Leitstrukturfindung (*lead discovery*) und Leitstrukturoptimierung (*lead optimization*) unterschieden. Die vorliegende Arbeit beinhaltet Beiträge zu beiden Bereichen.

Neben ausreichender **Affinität** eines Arzneistoffs zu seinem makromolekularen Rezeptor (engl.: *Target*) ist die **Selektivität** gegenüber anderen potentiellen Targets (sog. *Antitargets*) von großer Bedeutung. In vielen Fällen führt die Interaktion des Arzneistoffes mit Rezeptoren, die nicht die eigentliche, erwünschte Wirkung der Substanz vermitteln, zu sog. *unerwünschten Arzneimittelwirkungen* (Abk. UAW, oft auch etwas unklar als *Nebenwirkungen* bezeichnet).

Beispiele für solche *pharmakodynamischen* Nebenwirkungen sind Herzrhythmusstörungen bei Einnahme trizyklischer Antidepressiva, die durch deren antagonistische Wirkung an Muskarinrezeptoren und Natriumkanalblockade hervorgerufen werden, die bronchokonstriktorische Komponente von β -Blockern, die auf Antagonismus von β_2 -Rezeptoren zurückzuführen ist, oder die Knochenmarkstoxizität als ernste Nebenwirkung von Chloramphenicol, das nicht nur die Translation bakterieller, sondern auch mitochondrialer Proteine hemmt. Dennoch ist zu bemerken, dass bei einigen Arzneistoffen der Mangel an Selektivität für die Hauptwirkung entscheidend ist, so bei den eben erwähnten trizyklischen Antidepressiva, deren Hemmung verschiedener Monoamintransporter für die angstlösende, stimmungsaufhellende Wirkung verantwortlich ist. Weitere prominente Vertreter solcher sog. *dirty drugs* sind atypische Neuroleptika: Sie antagonisieren neben den für die antipsychotische Wirkung verantwortlichen D_2 -Rezeptoren eine Reihe weiterer Neurorezeptoren (D_1 , D_4 , α_1 , 5-HT_{2A}), was als eine mögliche Erklärung für das Fehlen der gefürchteten extrapyramidalmotorischen Störungen (EPS) herkömmlicher Neuroleptika diskutiert wird [Müller, 2002]. Ein Antitarget, das von vielen, strukturell verschiedenartigen Wirkstoffklassen inhibiert wird und dadurch zu kardialen Nebenwirkungen führt, ist der hERG-Kanal, ein spannungsgesteuerter Kaliumkanal, dessen Hemmung zu anormalen Phänomenen im EKG und zu Arrhythmien führen kann [Vandenberg et al., 2001].

Ein weiterer Bereich als Folge mangelnder Selektivität betrifft die *Pharmakokinetik*; es handelt sich dabei um Interaktionen mit Systemen der Arzneistoffaufnahme, -verteilung, -metabolisierung und -exkretion. Von besonderer Bedeutung hierbei sind die Cytochrom P450-Enzyme. Werden sie inhibiert, kann dies zu einer Wirkungsverstärkung oder -verlängerung anderer Xenobiotika führen. Die Induktion von CYP450 hingegen kann zur Verminderung oder Verkürzung der Wirkung führen. Ein prominentes Beispiel ist die Induktion durch Phenobarbital, die bei Patienten unter antikoagulatorischer Cumarintherapie thromboembolische Komplikationen verursachen kann. Von besonderer Bedeutung für die (Neben)Wirkungen ZNS-aktiver Wirkstoffe ist die Interferenz mit dem P-Glykoprotein (P-gp), das u.a. den Transport von Wirkstoffen aus dem ZNS heraus über die Blut-Hirn-Schranke bewerkstelligt und somit zur Unwirksamkeit verschiedener Substanzen im ZNS führt. Auch bei der Tumorthherapie werden Zytostatika nicht selten aufgrund der Aktivität des P-gp aus der Tumorzelle ausgeschleust und damit unwirksam. Diesen Eigenschaften hat das für P-gp kodierende Gen auch den Namen MDR1- (*Multiple Drug Resistance*) Gen zu verdanken.

Diese vorgestellten Beispiele machen die Signifikanz von Selektivitätsaspekten in der Wirkstoffentwicklung, insbesondere in der Phase der Leitstruktoptimierung, deutlich. Es ist daher erstrebenswert, derartige Überlegungen möglichst früh in den Entwicklungsprozess einzubeziehen. Die Methoden des *Molecular Modelling* stellen hierbei eine Möglichkeit dar, um mit Hilfe computerbasierter Verfahren Vorhersagen von Selektivitäten zu treffen. In der vorliegenden Arbeit werden Computermodelle entwickelt und angewendet, die geeignet sind, um Affinität und Selektivität von Enzyminhibitoren vorherzusagen. Dazu werden insbesondere **QSAR**-Methoden (*Quantitative Structure Activity Relationship*) verwendet.

In Kapitel 2 wird eine Einführung in die theoretischen Grundlagen der verwendeten Computermethoden sowie ein kurzer Überblick von Beispielen aus der Literatur zu diesem Themenbereich gegeben. Kap. 3 beschreibt strukturelle und funktionale Aspekte der zahlreichen Isoenzyme der Carboanhydrase (CA), die als Modellsystem für diese Studien dienen. Kap. 4 schildert die Generierung und Validierung der 3D-QSAR-Modelle zur Affinitäts- und Selektivitätsvorsage an den Isoenzymen I, II und IV der Carboanhydrase sowie einen Vergleich der verschiedenen Verfahren in Bezug auf die statistischen Parameter und die Konturkarten. Um 3D-QSAR-Modelle zum Screening größerer Datenbanken zu verwenden, wurde ein Protokoll zur automatischen Ligandüberlagerung angewandt und die Ergebnisse mit alignmentfreien Methoden (1D-, 2D-QSAR) verglichen (Kap. 5). Kapitel 6 beschreibt die Anwendung dieses Alignment-

protokolls und der entwickelten Modelle zur Bewertung einer mehrere tausend Moleküle umfassenden virtuellen Bibliothek.

1.2 Virtuelles Screening zur Leitstrukturfindung

Der letzte und thematisch abgesetzte Teil der Arbeit (Kap. 7) beschäftigt sich mit einer Technik der Leitstrukturfindung, dem sog. **virtuellen Screening** [Klebe, 2006]. Virtuelles Screening (VS) ist eine kostengünstige Alternative oder Ergänzung zum – meist industriell durchgeführten – *High-Throughput Screening* (HTS). Beim HTS werden mit Hilfe von Robotern Substanzkollektionen in Größenordnungen von ca. 500000 bis 1 Mio. Verbindungen gegen Rezeptoren oder ganze Zellen auf biologische Aktivität gemessen. Nachteile dabei sind Kosten- und Zeitaufwand, verhältnismäßig hohe falsch-positive oder falsch-negative Raten und die Tatsache, dass die Substanzen auch bereits synthetisiert sein müssen. Beim VS hingegen kann jedes beliebige (auch fiktive) Molekül *in silico* bezüglich seiner Wahrscheinlichkeit, die erwünschte Wirkung hervorzurufen, abgeschätzt werden. Dies kann aufgrund von *Pharmakophormodellen*, Ähnlichkeitssuchen, QSAR-Modellen, Docking und Scoring etc. erfolgen. In letzterem Fall ist es notwendig, die Struktur des Rezeptors zu kennen. Der besondere Vorteil von VS-Verfahren liegt in der Geschwindigkeit und der Kostenersparnis. Es kann auch von Vorteil sein, VS **komplementär** zum HTS einzusetzen, entweder um falsch-negative Moleküle zu „retten“ oder zum Design *fokussierter* Substanzbibliotheken, die anschließend genauer und mit geringerem Durchsatz biochemisch vermessen werden können [Schnecke & Boström, 2006]. In der Literatur sind zahlreiche Fälle von erfolgreich durchgeführten virtuellen Screenings zu finden [Kubinyi, 2006].

Im Rahmen der vorliegenden Dissertation wurde ein virtuelles Screening an Peptid-deformylasen als Target durchgeführt. Zum Einsatz kam eine Kombination aus einem strukturbasierten Pharmakophormodell in Verbindung mit Docking und Scoring zur Auswahl kommerziell erhältlicher Verbindungen.

2 Selektivitätsvorhersage mittels chemometrischer Methoden

2.1 Ausschließliche Verwendung von Ligandinformation: 3D QSAR-Verfahren (CoMFA und CoMSIA)

Quantitative Struktur-Wirkungsbeziehungen (engl.: *Quantitative Structure Activity Relationships*, **QSAR**) sind mathematische Modelle, die den Zusammenhang zwischen chemischer Struktur und biologischer Aktivität von Molekülen erstellen. Die chemische Struktur muss dazu in Form von Zahlenwerten kodiert werden, den sog. Deskriptoren, die entweder experimentell gemessene oder theoretisch berechnete Größen darstellen, in jedem Fall aber eine Funktion der chemischen Struktur sein müssen. Die vermutlich ältesten und bekanntesten QSAR-Deskriptoren, die seit der Etablierung dieses Forschungsgebietes durch Corwin Hansch 1964 breite Anwendung gefunden haben [Hansch & Fujita, 1964; Kubinyi, 1993b], sind die Hammett-Konstante σ von aromatischen Substituenten und der Octanol-Wasser-Verteilungskoeffizient P einer Substanz. Als Größe für die biologische Aktivität wird häufig der negative dekadische Logarithmus des IC_{50} - oder K_i -Wertes, gewonnen aus einem Assay, verwendet, aber auch LD_{50} -Werte für toxikologische QSARs oder chromatographische Retentionszeiten (dann oft als *Quantitative Structure Property Relationship*, QSPR, bezeichnet) können beispielsweise verwendet werden. Um eine Korrelation zwischen Moleküldeskriptoren (auch unabhängige Variable, Prädiktor) und der Zielvariablen (abhängige Variable) zu erhalten, werden verschiedene Methoden der Statistik oder des Maschinellen Lernens verwendet, z. B. Multiple Lineare Regression (MLR [Kubinyi, 1993b], auch *Ordinary Least Squares* – OLS), *Principal Components Regression* (PCR), *Partial Least Squares* (PLS) Verfahren [Geladi & Kowalski, 1986], Künstliche Neuronale Netze (ANN) [Zupan & Gasteiger, 1999], *Bayes*-Klassifikatoren, *Support Vector Machines* (SVM) [Witten & Frank, 2005] u.v.a.m.

In der vorliegenden Arbeit lag der Schwerpunkt auf 3D-QSAR-Methoden [Kubinyi, 1993a], auf deren theoretische Grundlagen daher im Folgenden genauer eingegangen

werden soll, auch um das im nächsten Abschnitt behandelte AFMoC-Verfahren besser verstehen zu können. Bei den 3D-QSAR-Methoden, deren verbreitetste das **CoMFA**- und **CoMSIA**-Verfahren darstellen, müssen von den zu untersuchenden Liganden zunächst Konformationen erzeugt und diese anschließend räumlich überlagert werden (sog. *Alignment*). Auf die dazu verwendeten Methoden wird in Kap. 2.4 näher eingegangen. Die überlagerten Liganden werden dann in einer Gitterbox platziert, wobei auf jedem Gitterpunkt ein Sondenatom lokalisiert ist. Je nach Verfahren werden dann mit Hilfe einer abstandsabhängigen Funktion Wechselwirkungsterme zwischen dem Sondenatom und den Molekülatomen berechnet und aufsummiert. Der für jedes Sondenatom berechnete Wert wird in einem Zeilenvektor abgelegt. Ordnet man die auf diese Weise für jedes Molekül erhaltenen Vektoren untereinander an, so erhält man – bei m Molekülen und n Wechselwirkungstermen – eine $m \times n$ -Deskriptormatrix (X-Matrix). In der Regel ist jedem Molekül ein Aktivitätswert zugeordnet, der in der zugehörigen Y-Matrix abgelegt wird (Abb. 2.1).

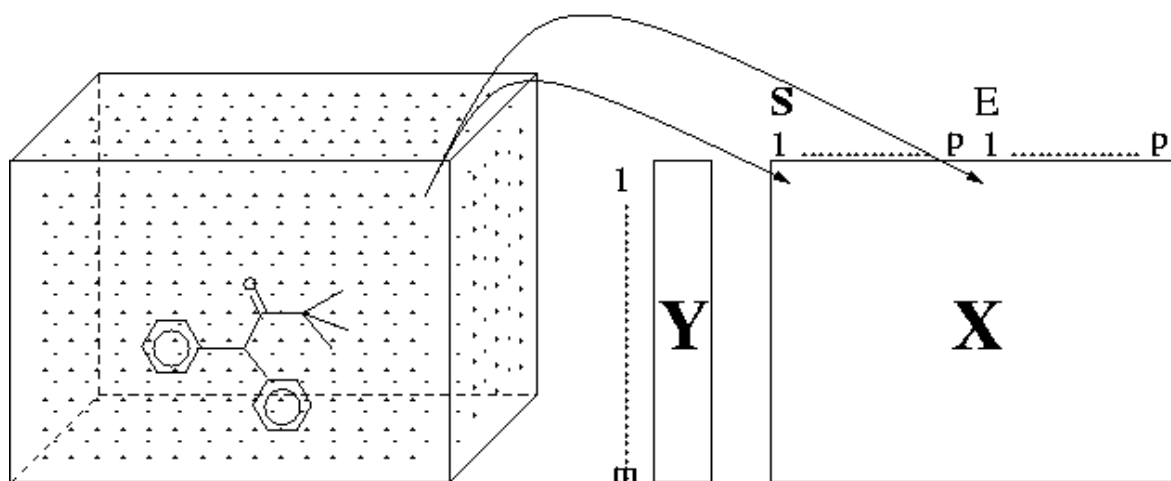


Abbildung 2.1: Schematische Darstellung des Aufbaus von X- und Y-Matrix für 3D-QSAR-Studien. Die Abbildung zeigt eine Matrix mit m Molekülen am Beispiel einer CoMFA-Analyse, bei der jeweils p Energiewerte für sterische (S) und elektrostatische (E) Wechselwirkungen abgelegt werden (Abb. aus [Hillebrecht, 2005]).

Der Hauptunterschied zwischen CoMFA (*Comparative Molecular Field Analysis*) [Cramer, 1988] und CoMSIA (*Comparative Molecular Similarity Indices Analysis*) [Klebe et al., 1994] besteht in der funktionalen Form, die zur Berechnung der Wechselwirkungsterme verwendet wird. Bei CoMFA wird zur Beschreibung sterischer Wechselwirkungen S_p das LENNARD-JONES-Potential benutzt, während elektrostatische Wechselwirkungen E_p mit Hilfe des COULOMB-Gesetzes modelliert werden.

$$S_p = \sum_{i=1}^n \frac{A_{pi}}{r_{pi}^{12}} - \frac{B_{pi}}{r_{pi}^6} \quad (2.1)$$

$$E_p = \sum_{i=1}^n \frac{q_p q_i}{4\pi r_{pi}^2 \epsilon_0 \epsilon_r} \quad (2.2)$$

Darin bedeuten S_p und E_p das LENNARD-JONES- bzw. COULOMB-Potential am Punkt p , r_{pi} ist der Abstand zwischen dem Gitterpunkt p und dem Molekülatom i (das Molekül besitzt n Atome), A und B sind empirische Größen, die von den Eigenschaften des Sonden- und Molekülatoms abhängen, q_p und q_i sind die Ladungen des Sonden- und des Molekülatoms, ϵ_0 ist die Permittivität des Vakuums und ϵ_r ist die Dielektrizitätskonstante des jeweiligen Mediums. Aus der funktionellen Form des LENNARD-JONES-Potentials ergibt sich bereits ein gravierender Nachteil des CoMFA-Verfahrens: Durch den steilen Anstieg bedingt, strebt der Wert von S_p in der Nähe des Molekülatomkerns gegen unendlich. Um Probleme bei der nachfolgenden PLS-Analyse zu vermeiden, muss ein willkürlicher *cut-off*-Wert (standardmäßig 30 kcal/mol) gesetzt werden, der nicht überschritten werden darf (Abb. 2.2, links). Als Konsequenz umschließen CoMFA-Konturen niemals die Molekülareale, durch die sie hervorgerufen wurden, sondern liegen außerhalb, aber dicht daneben. Zusätzlich bewirkt der steile Funktionsverlauf, dass die Konturen meist ein eckiges, zerrissenes Aussehen besitzen, was die Interpretierbarkeit u.U. erschwert. Außerdem werden alle aktivitätsbeeinflussenden Faktoren durch nur zwei Eigenschaften, sterische und elektrostatische, zusammengefasst, so dass nur eine relativ wenig differenzierte Interpretation möglich ist.

Aufgrund dieser Unzulänglichkeiten entwickelten Klebe et al. das CoMSIA-Verfahren [Klebe et al., 1994; Klebe & Abraham, 1999]. Hierbei werden sog. **Ähnlichkeitsindizes** als 3D-Deskriptoren verwendet. Sie werden nicht durch reale physikalische Größen, sondern durch Gaussfunktionen berechnet. Der Ähnlichkeitsindex $A_k^q(j)$ zwischen einer Sonde am Gitterpunkt q und einem Molekül j für eine bestimmte physikochemische Eigenschaft k ist definiert als

$$A_k^q(j) = - \sum_{i=1}^n w_{Probe,k} w_{ik} e^{-\alpha r_{iq}^2} \quad (2.3)$$

Dabei ist n die Anzahl der Atome des Liganden j , $w_{Probe,k}$ ist der Wert der physikochemischen Eigenschaft k des Sondenatoms, w_{ik} der Wert der physikochemischen

Eigenschaft k des Atoms i des Liganden, r_{iq} der Abstand des Atoms i zum Gitterpunkt q , und α ist der Abschwächungsfaktor. Dieser Abschwächungsfaktor bestimmt, wie schnell die Gauss-Funktion abfällt; bei einem niedrigen Wert für α fällt die Funktion schnell ab, und lokale Ähnlichkeiten spielen eine größere Rolle bei der Berechnung der Ähnlichkeitsindizes. Ein geringer Wert für α hingegen führt zur Berücksichtigung globaler Ähnlichkeiten.

Die Verwendung von Gaussfunktionen hat den Vorteil, dass diese einen relativ weichen Funktionsverlauf besitzen; daher erscheinen CoMSIA-Konturen meist glatter und runder. Die statistischen Parameter aus einer PLS-Analyse von CoMSIA-Deskriptoren sind wesentlich robuster und unempfindlicher gegen Translationen und Rotationen des Gitterkastens als bei CoMFA-Deskriptoren [Böhm, 1998]. Ferner streben Gaussfunktionen am Atomkern gegen einen endlichen Maximalwert, so dass das Setzen willkürlicher cut-off-Werte hierbei entfällt. Ein graphischer Vergleich der Verläufe der für CoMFA- und CoMSIA-Verfahren verwendeten Abstandsfunktionen ist in Abb. 2.2 gezeigt.

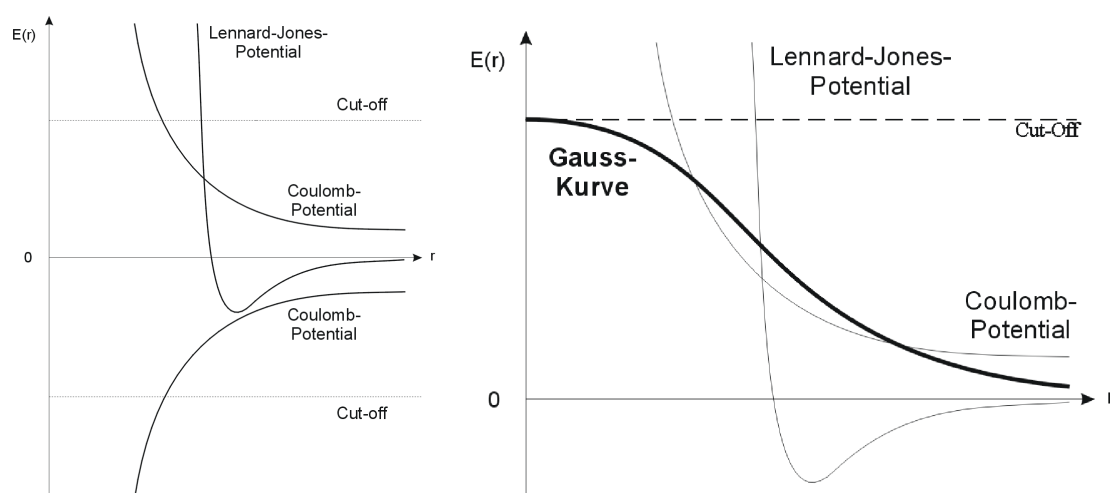


Abbildung 2.2: Vergleich der funktionellen Verläufe von Lennard-Jones(12,6)- und Coulomb-Potential (links) mit dem der Ähnlichkeitsindizes (rechts). Die Steilheit der CoMFA-Funktionen macht die Definition eines willkürlichen *cut-off*-Wertes (gestrichelte Linie) notwendig, während die wesentlich weicher verlaufende Gaussfunktion bei $r = 0$ gegen einen endlichen Wert strebt.

Beim CoMSIA-Verfahren werden Ähnlichkeitsindizes für fünf physikochemische Eigenschaften verwendet, sterische, elektrostatische, hydrophobe, Wasserstoffbrückendonator- und Wasserstoffbrückenakzeptoreigenschaften. Durch diese Aufteilung werden die wesentlichen Faktoren, welche Protein-Ligand-Wechselwirkungen bestimmen, abgedeckt.

Dies ist ein weiterer Vorteil gegenüber dem CoMFA-Verfahren, bei dem hydrophobe Effekte oft durch die sterischen und Wasserstoffbrückeneigenschaften durch die elektrostatischen Eigenschaften abgedeckt werden. Die Deskriptoren werden bei beiden Verfahren mit Hilfe der PLS-Analyse mit den biologischen Aktivitäten korreliert. Dadurch, dass wenigen Aktivitätsdaten (typischerweise 20 bis 100) sehr viele (mehrere 1000) Deskriptoren gegenüberstehen, entsteht ein stark unterbestimmtes System, das mit gewöhnlicher multipler linearer Regression (MLR) nicht lösbar ist. Zudem sind die Deskriptoren bei 3D-QSAR-Verfahren meist hoch korreliert, was ebenfalls ein Problem für die MLR darstellt. Die PLS-Analyse liefert schließlich eine QSAR-Gleichung der Form

$$y = a_1S_1 + a_2S_2 + \dots + a_pS_p + b_1E_1 + b_2E_2 + \dots + b_pE_p + k \quad (2.4)$$

im Falle des CoMFA-Verfahrens, wobei S_i und E_i die entsprechenden sterischen bzw. elektrostatischen Wechselwirkungsenergien an den p Gitterpunkten sind, a_i und b_i sind die entsprechenden Koeffizienten, und k ist ein konstanter Summand. Solche Gleichungen können dann herangezogen werden, um die Affinität neu synthetisierter Verbindungen zu berechnen. Die Koeffizienten können ferner auf die jeweiligen Gitterpunkte zurückprojiziert werden; konturiert man anschließend diejenigen Bereiche, innerhalb derer die Koeffizienten einen bestimmten Mindestwert (sog. Konturlevel) unter- oder überschreiten, so erhält man (Iso)Konturkarten. Sie zeigen Bereiche an, deren Besetzung mit Molekülteilen einer bestimmten Eigenschaft die Aktivität erhöhen oder erniedrigen. Solche Konturplots können daher das Verständnis der zugrunde liegenden Struktur-Wirkungsbeziehung erhellen oder dazu dienen, Vorschläge zum Design verbesserter Liganden zu erhalten.

2.2 Gleichzeitige Verwendung von Protein- und Ligandinformation: AFMoC

Die oben beschriebenen 3D-QSAR-Verfahren benötigen keinerlei Information über den Rezeptor, an den die untersuchten Liganden binden. Dies kann als großer Vorteil angesehen werden, da für viele pharmazeutisch relevante Targets keine Kristallstrukturen vorhanden sind, z.B. bei G-Protein gekoppelten Rezeptoren (GPCRs) oder Ionenkanälen. Hier stellen QSAR-Verfahren oft die einzige Möglichkeit zur Anwendung

des rationalen Wirkstoffdesigns dar. Ist eine Kristallstruktur des Rezeptors aber vorhanden, so kann diese wertvolle Informationen über affinitätsbestimmende Aspekte liefern. So genannte *Scoring*-Funktionen können verwendet werden, um Affinitäten abzuschätzen, allerdings eignen sich nicht alle Scoring-Funktionen dazu. Andere sind eher geeignet, aus verschiedenen Geometrien eines Liganden in der Bindetasche diejenige herauszufiltern, die der tatsächlich in einer Kristallstruktur beobachteten am nächsten kommt, meist in Verbindung mit *Docking*-Programmen, die die Geometrien erzeugen. Da Scoringfunktionen i.A. von einer Vielzahl verschiedenartiger Proteinfamilien abgeleitet werden, sind sie zwar allgemeiner anwendbar, besitzen aber oft für ein spezielles Target nur unzureichende Genauigkeit.

Das **AFMoC**-Verfahren (*Adaptation of Fields for Molecular Comparison*) [Gohlke & Klebe, 2002b; Gohlke, 2000] versucht, durch Einbeziehung von Proteininformation zu ermöglichen, mit weniger Liganden als bei herkömmlichen QSAR-Studien (oft mindestens 20) auszukommen. Gleichzeitig soll umgekehrt die Ligandinformation benutzt werden, um eine allgemein gültige Scoring-Funktion auf das jeweilige Protein anzupassen und dadurch eine höhere Genauigkeit bei der Affinitätsabschätzung zu erzielen. Im Folgenden sollen die Details des Verfahrens erörtert werden.

Wie bei den zuvor besprochenen Methoden muss zunächst ein relevantes Alignment der untersuchten Liganden erzeugt werden. Allerdings muss dieses beim AFMoC-Verfahren in der Proteinbindetasche geschehen, wie in der weiteren Ausführung deutlich wird. Die bekannte Gitterbox wird so platziert, dass sie die Bindetasche und die Liganden ausreichend umschließt. An jedem Gitterpunkt wird ein Sondenatom platziert, wobei alle der 17 Atomtypen, welche bei der Ableitung der Scoring-Funktion DrugScore berücksichtigt wurden, verwendet werden können. DrugScore [Gohlke et al., 2000a,b; Gohlke, 2000; Gohlke & Klebe, 2001, 2002a] ist eine wissensbasierte Scoring-Funktion; dieser Ansatz basiert auf einer Auszählung der statistischen Häufigkeiten, mit denen zwei bestimmte Atomtypen in Kristallstrukturen mit einer Distanz von 2–6 Å beieinander zu liegen kommen. Unter Einbeziehung eines Referenzzustandes, der die Wahrscheinlichkeiten für das Auftreten eines beliebigen Atom-Atom-Kontaktes beschreibt, können entsprechende „Pseudoenergien“ oder statistische Präferenzen, die sog. **Paarpotentiale**, errechnet werden. Ein im Vergleich zum Referenzzustand häufiges Auftreten eines bestimmten Atom-Atom-Kontaktes entspricht hierbei einer niedrigen „Energie“ und umgekehrt. Dieses Konzept wurde anfangs auf Proteinkristallstrukturen aus der PDB angewendet. Aufgrund der besseren Datenlage und höheren Auflösung niedermolekularer Kristallstrukturen wurden analoge Potentiale später auch von CSD-Strukturen

abgeleitet, was zu einer Verbesserung der Performance führte [Velec et al., 2005]. Die Stärke von DrugScore liegt insbesondere in der Identifizierung von Bindungsgeometrien, wie sie in Kristallstrukturen beobachtet werden. Außer zum Rescoring wurde DrugScore auch als Zielfunktion für das Docking verwendet [Sottriffer et al., 2002]. Ferner können die Paarpotentiale dazu benutzt werden, Regionen, die besonders günstig für einen speziellen Atomtyp in einer Bindetasche sind (sog. *hot spots*), zu identifizieren. Die Anwendbarkeit zur Affinitätsabschätzung ist ebenfalls demonstriert worden [Gohlke et al., 2000b; Gohlke & Klebe, 2002a].

Im Sinne einer *hot spots* Analyse werden beim AFMoC-Ansatz zunächst „Wechselwirkungsenergien“ jedes Sondenatoms mit allen Proteinatomen berechnet. Dadurch wird die Proteininformation ausgenutzt. Diese Potentialwerte werden anschließend mit Gewichtungsfaktoren B multipliziert, die eine gaussartige Abstandsabhängigkeit von den jeweiligen Atomtypen aufweisen:

$$B_{g,l} = \frac{e^{-\frac{r_{g,l}^2}{2\sigma^2}}}{\sum_{i=1}^n e^{-\frac{r_{g,i}^2}{2\sigma^2}}} \quad (2.5)$$

Hierin ist $B_{g,l}$ der Gewichtungsfaktor am Gitterpunkt g , ausgehend von einem Ligandatom l , $r_{g,l}$ ist der Abstand zwischen g und l , und σ ist die Breite der Gaussfunktion. Die Summe im Nenner läuft über alle n Gitterpunkte und bewirkt eine Normierung des Gewichtungsfaktors.

Auf diese Weise erhält man – analog zu den zuvor beschriebenen 3D-QSAR-Verfahren – für jeden Ligand einen Vektor, dessen Länge abhängt von der Zahl der Gitterpunkte und den betrachteten Atomtypen. Die Affinitätsbeiträge von Atomtypen, die in mindestens einem der Liganden enthalten sind, aber nicht in der anschließenden PLS-Analyse verwendet werden, werden durch die rohen DrugScore-Potentiale über einen empirisch ermittelten Skalierungsfaktor abgeschätzt. Abb. 2.3 stellt die Berechnung der „Wechselwirkungsfelder“ anschaulich dar.

Zunächst werden für jeden Gitterpunkt in der Proteinbindetasche die unveränderten DrugScore *hot spots* berechnet (Abb. 2.3, links). Durch Multiplikation mit den abstandsabhängigen Gewichtungsfaktoren, die von auf den entsprechenden Ligandatom zentrierten Gaussfunktionen abgeleitet wurden (Mitte), werden die entsprechenden Wechselwirkungsfelder dort abgeschwächt, wo keine Ligandatome vorkommen (Abb.

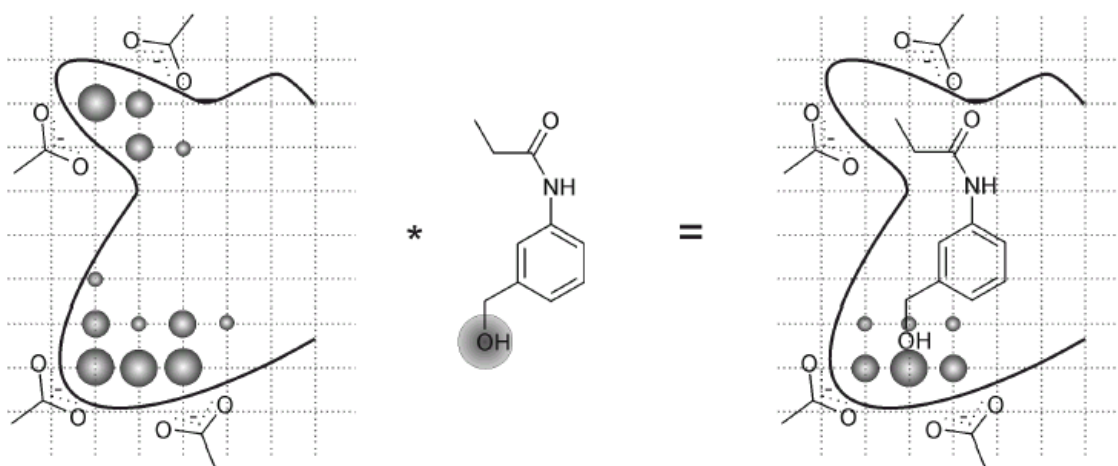


Abbildung 2.3: Schematische Darstellung der Berechnung von Wechselwirkungsfeldern zur Ableitung von AFMoC-Deskriptoren. Im linken Teil der Graphik ist eine stilisierte Proteinbindetasche zu sehen, in die ein regelmäßiges Gitter eingeschrieben worden ist. An jedem Gitterpunkt werden DrugScore-Potentiale, hier für den Atomtyp O.3, berechnet (*hot spots*). Für diesen Atomtyp wird eine Gaussfunktion auf entsprechenden Ligandatomen zentriert (hier auf dem O einer OH-Gruppe); von der Gaussfunktion werden Gewichtungsfaktoren abgeleitet, mit denen die ursprünglichen DrugScore-Potentiale multipliziert werden (Abb. aus [Gohlke & Klebe, 2002b]).

2.3, rechts oben). An Stellen, an denen der entsprechende Atomtyp hingegen beobachtet wird, werden die ursprünglichen Potentialwerte verstärkt (rechts unten). Auf diese Weise wird die generische Proteininformation mit spezifischer Ligandinformation kombiniert. Wie bei den anderen 3D-QSAR-Verfahren wird eine Korrelation der so erhaltenen Deskriptormatrix mit den zugehörigen Affinitätswerten durch eine PLS-Analyse bewerkstelligt.

Eine Besonderheit des AFMoC-Verfahrens besteht darin, dass zur Vorhersage des pK_i -Wertes einer Verbindung nicht ausschließlich die Koeffizienten der QSAR-Gleichung herangezogen werden: Zunächst wird der Gesamt- pK_i -Wert pK_i eines Liganden als Summe zweier Anteile, pK'_i und pK''_i , ausgedrückt. Dabei ist pK'_i der durch die in der PLS-Analyse betrachteten Atomtypen abgedeckte Anteil, pK''_i ist der Anteil, der durch die originalen DrugScore-Potentiale mit Hilfe des empirischen Skalierungsfaktors c_S^{Pair} berechnet wird, für Atomtypen, die nicht in die PLS-Analyse aufgenommen wurden:

$$pK_i = pK'_i + pK''_i \quad (2.6)$$

Die Berechnung des Anteils pK'_i ergibt sich dann nach folgender Gleichung:

$$pK'_i = \Theta < pK'_i > + \sum_{g \in G} \sum_{l \in L: T(l) \in T'} (\Theta c_{g,T(l)} \frac{\Delta W_{g,T(l)} - < \Delta W_{g,T(l)} >}{\Delta \bar{W}_{g,T(l)}} + (1 - \Theta) c_S^{Pair} \Delta W_{g,T(l)}) \quad (2.7)$$

Dabei ist

- c_S^{Pair} der Scaling Faktor der DrugScore-Funktion
- $c_{g,t}$ der für den Atomtyp t und den Gitterpunkt g berechnete QSAR-Koeffizient
- l ein Ligandenatom aus der Gesamtheit L
- g ein Gitterpunkt der Gesamtheit G
- T die Gesamtheit aller Atomtypen (die DrugScore unterscheidet)
- T' diejenige Teilmenge von T , die zur Erstellung des AFMoC-Models herangezogen wurde
- Θ der sog. *mixing factor*
- $< pK'_i >$ der Mittelwert der pK'_i -Werte der Trainingsverbindungen
- $< \Delta W_{g,T(l)} >$ der Mittelwert der DrugScore-Paarpotentialwerte
- $\Delta \bar{W}_{g,T(l)}$ die Standardabweichung von $\Delta W_{g,T(l)}$.

Der *mixing factor* Θ kann Werte zwischen 0 und 1 annehmen; bei Werten nahe 0 werden hauptsächlich die ursprünglichen, unspezifischen DrugScore-Potentiale zur Affinitätsabschätzung verwendet. Diese Einstellung ist sinnvoll, wenn nur wenig Ligandinformation zur Verfügung steht. Je mehr Liganden mit bekannter Affinität allerdings bekannt sind, desto näher sollte Θ an 1 gewählt werden. Dann werden bevorzugt die adaptierten Wechselwirkungsfelder verwendet. Auf diese Weise kann der Benutzer einstellen, wie groß er die Anteile von Protein- und Ligandinformation gewichten möchte.

2.3 Verwendung von 3D-QSAR-Verfahren zur Vorhersage von Selektivität

Die in Kap. 2.1 und 2.2 vorgestellten 3D-QSAR-Verfahren können prinzipiell zur Vorhersage beliebiger, geeignet skalierteter Aktivitäts- oder Eigenschaftsvariablen herangezogen werden. In der Mehrzahl der Fälle sind dies Affinitäten zu Rezeptoren, die von Interesse sind, ausgedrückt als pK_i - oder pIC_{50} -Werte. Ebenso ist es aber auch möglich, in analoger Weise Selektivitäten in Bezug auf zwei Rezeptoren, A und B, vorherzusagen.

Hierbei können prinzipiell zwei Wege beschritten werden:

1. Ableitung von zwei separaten Modellen, eines für Rezeptor A und eines für Rezeptor B, zur Vorhersage von $pK_i(A)$ und $pK_i(B)$
2. Ableitung eines Modells, das auf die Differenz $pK_i(A) - pK_i(B)$ ($\Delta pK_i(A - B)$) trainiert wird.

Bei der letzteren Variante ist zu beachten, dass die unabhängige Variable ΔpK_i aus zwei einzelnen Variablen zusammengesetzt ist. Nach den Regeln der Fehlerfortpflanzung bewirkt die Differenzbildung der einzelnen pK_i -Werte eine Aufsummierung der Fehler, mit denen sie jeweils behaftet sind. Bei diesem Ansatz wird eine Konturkarte pro Deskriptorfeld erhalten, wobei die Konturen Bereiche markieren, an denen eine bestimmte physikochemische Eigenschaft die Selektivität bezüglich des einen oder anderen Rezeptors erhöht. Ein Problem stellt hierbei allerdings die Wahl des Alignments dar, da man sich (relativ willkürlich) zwischen dem Alignment für Rezeptor A oder B zu entscheiden hat.

Im Falle des ersten Ansatzes mit zwei separaten Modellen können theoretisch auch zwei getrennte Alignments und – sofern das AFMoC-Verfahren angewendet wird – auch beide Rezeptorstrukturen A und B einbezogen werden. Man erhält bei diesem Ansatz zwei einzelne Konturkarten, die jeweils getrennt Aufschluss über affinitätssteigernde und -vermindernde Einflüsse geben. Selektivitätsaussagen lassen sich daraus durch parallele Interpretation und visuellen Vergleich einander entsprechender Regionen erhalten. Alternativ ist es aber auch möglich, durch Subtraktion der QSAR-Koeffizienten einander entsprechender Gitterpunkte beide Konturkarten zu einer „Differenz-“ oder

„Selektivitätskarte“ zu kombinieren. In der vorliegenden Arbeit wurde diese Variante bevorzugt, da die Interpretation auf diese Weise erleichtert wird. Da die QSAR-Koeffizienten in diesem Fall aus zwei separat abgeleiteten Modellen stammen, ist es möglich, dass sie unterschiedliche Wertebereiche abdecken. Daher wurde neben der direkten Differenzbildung auch noch überprüft, ob eine Normierung der Koeffizienten auf einen Bereich von 0 bis 1 *vor* der Differenzbildung Unterschiede in den Konturkarten bewirkt. Ein weiteres Problem, welches bei der Variante der Differenzbildung der Koeffizienten auftreten kann, ist, dass – bei Verwendung unterschiedlicher Alignments – unsignifikante Alignmentunterschiede als selektivitätsbestimmende Faktoren durch die PLS-Analyse hervorgehoben werden. So kann eine leichte und unbedeutsame, eher zufällige Verschiebung einer funktionellen Gruppe eines Liganden (hervorgerufen etwa durch leichte Ungenauigkeiten beim Docking oder Minimieren) mit Differenzen in den Affinitäten korreliert werden und entsprechende Konturen hervorrufen. Daher kann es selbst im Fall des ersten Ansatzes geboten sein, nur ein Alignment zu verwenden, um nicht ein unnötiges „Hintergrundrauschen“ (*noise*) in der multivariaten Datenstruktur zu verstärken.

2.4 Alignmentverfahren

Der schwierigste und arbeitsintensivste Schritt bei der Ableitung von 3D-QSAR-Modellen stellt das *Alignment*, also die Erzeugung von Ligandkonformationen und deren räumliche Überlagerung, dar. Idealerweise würden hierzu die bioaktiven Konformationen verwendet, wie sie in Kristallstrukturen von Protein-Ligand-Komplexen zu finden sind. Allerdings steht fast niemals für alle Liganden der untersuchten Serie diese Information zur Verfügung. Es konnte sogar gezeigt werden, dass ein in sich konsistentes, aber artifizielles Alignment, welches nicht den bioaktiven Konformationen entspricht, bessere Ergebnisse liefern kann als ein auf Kristallstrukturen basierendes [Klebe & Abraham, 1993].

Verschiedene Ansätze wurden entwickelt, um zu einem Alignment zu gelangen [Klebe, 1993; Lemmen & Lengauer, 2000]: In der einen Gruppe von Methoden wird die Proteinstruktur des Targets verwendet, so bei **Docking**-Verfahren oder **manueller** Platzierung der Liganden in der Bindetsche und anschließender Energieminimierung. Andere Verfahren verwenden die Erkennung von Molekülteilen ähnlicher **atomarer** Struktur (*maximal common substructure search*, *MCSS* [Handschuh et al., 1998], *mul-*

tiptle atom fit) oder funktioneller Gruppen (z. B. DISCO, CATALYST, GASP). Eine weitere Verallgemeinerung erfahren diese Konzepte in den **feldbasierten** Verfahren, die auch den 3D-QSAR-Methoden in der Beschreibung der physikochemischen Eigenschaften im Hinblick auf die molekulare Erkennung durch den Rezeptor am ähnlichsten sind. Prototyp dieses Ansatzes ist SEAL [Kearsley & Smith, 1990], das die physikochemischen Eigenschaften in Form von Gaussfunktionen übersetzt. Weitere feldbasierte Verfahren mit ähnlichem Konzept sind der **Field-Fit**-Ansatz, TORSEAL und FlexS, mit unterschiedlicher Berücksichtigung der Molekülflexibilität. Im Folgenden sollen nur kurz die Methoden beschrieben werden, die im Rahmen dieser Arbeit für 3D-QSAR-Studien verwendet wurden.

2.4.1 Manuelle Platzierung und anschließende Minimierung in der Bindetasche

Die Liganden wurden manuell in der Bindetasche platziert, wobei ein Komplex mit möglichst ähnlichem Inhibitor als Templat benutzt wurde. Besonders die Orientierung von konservierten Molekülteilen und Ankergruppen wurde durch einen paarweisen RMSD-Atom-Fit möglichst analog zum Templat eingestellt. Das im Programmpaket MOLOC implementierte MAB-Kraftfeld [Gerber & Müller, 1995] wurde anschließend zur Minimierung verwendet. Dabei wurden alle im Trainingsdatensatz gemeinsame Ankergruppen fixiert. Dadurch soll gewährleistet werden, dass minimale relative Verschiebungen der Liganden zueinander unnötiges Hintergrundrauschen („*noise*“) hervorrufen und dadurch Artefakte in den Konturplots entstehen.

Nachteil dieser Methode ist die zum Teil subjektive Natur der Auswahl, mit der die Startgeometrie bestimmt wird. Wird diese falsch getroffen, kann auch der folgende Minimierungsschritt dies nicht korrigieren, da er nur das nächste lokale Minimum finden wird. Der Hauptnachteil besteht in dem extremen Zeit- und Arbeitsaufwand dieser Methode, die daher nur auf kleinere Datensätze, auch in Bezug auf die Vorhersage, beschränkt ist.

2.4.2 Docking: FlexX, AutoDock, GOLD

Als leicht automatisierbare Alternative zur manuellen Platzierung in der Bindetasche bietet sich das *Docking* an. Diese Methode wurde bereits häufig in der Literatur zum Alignment für 3D-QSAR-Studien verwendet (z. B. [Sippl, 2002; Tervo et al., 2004; Zhang et al., 2007b]). Der generelle Nachteil dabei ist, dass zwar die Erzeugung der Dockingposen vollautomatisch erfolgt, jedoch sind Scoring-Funktionen meist nicht zuverlässig genug, um unüberwacht die richtige Geometrie auszuwählen, so dass bei großen Screening-Szenarien das Ausmaß an erforderlicher Intervention durch den Benutzer erheblich werden kann. In der vorliegenden Arbeit wurden drei Dockingprogramme verwendet, um sie auf ihre Eignung zum Alignment zu testen, die im Folgenden kurz beschrieben werden. Die Ergebnisse werden nicht im Einzelnen beschrieben, da die Performance in allen Fällen unzureichend war und daher ligandbasierte Methoden verwendet wurden. GOLD und FlexX wurden ferner im Rahmen des virtuellen Screening (Kap. 7) eingesetzt.

FlexX

Das Dockingprogramm **FlexX** [Rarey et al., 1996, 1997] verwendet einen fragmentbasierten Ansatz in Kombination mit einem **inkrementellen** Aufbaualgorithmus zur Ligandplatzierung. Dazu wird das Molekül zunächst in kleinere Fragmente zerlegt und aus diesen ein geeignetes **Basisfragment** zur Initialisierung ausgewählt, welches möglichst viele gerichtete Wechselwirkungen mit dem Protein eingehen können sollte. Nach dessen Platzierung werden nacheinander die übrigen Fragmente angefügt, wobei die Flexibilität durch Auswahl von Torsionswinkeln, die häufig in Kristallstrukturen gefunden werden, modelliert wird (Torsionsdatenbank MIMUMBA [Klebe & Mietzner, 1994]). Durch den aufeinanderfolgenden Anbau der Fragmente bildet sich eine baumartige Suchstruktur, wobei, um die kombinatorische Explosion zu beherrschen, immer nur die k besten Lösungen weiterverfolgt werden (sog. *k-greedy Algorithmus*). Die Protein-Ligand-Wechselwirkungen werden durch sog. Interaktionsgeometrien modelliert, d.h. zueinander komplementären Partnern, die eine physikochemische Wechselwirkung eingehen können, werden bestimmte Winkel- und Abstandsregeln zugeordnet, um deren Beitrag zu entsprechenden Scoring-Termen zu bestimmen.

In FlexX sind verschiedene Scoring-Funktionen implementiert: Original-FlexX-Score, ähnlich der von Böhm entwickelten Scoring-Funktion [Böhm, 1994], ChemScore [Eldridge

et al., 1997], PLP [Gelhaar et al., 1995b,a] und ScreenScore [Stahl & Rarey, 2001]. Weiterhin existieren verschiedene Erweiterungen des Programms für spezielle Anwendungen: FlexX-Pharm [Hindle et al., 2002] erlaubt die Verwendung von Pharmakophorfiltern während und nach dem Docking. Zum beschleunigten Docking großer kombinatorischer Substanzbibliotheken wurde FlexX^C [Rarey & Lengauer, 2000] entwickelt und schließlich mit der FlexX-Pharm-Funktionalität verbunden [Gastreich et al., 2006]. Ein spezielles Modul zur Berücksichtigung mehrerer Proteinkonformationen zur Simulation von *induced-fit* Phänomenen ist FlexE [Claussen et al., 2001].

AutoDock

AutoDock [Goodsell & Olson, 1990; Morris et al., 1996, 1998] verwendet zur Beschreibung von Protein-Ligand-Wechselwirkungen keine an funktionelle Gruppen gebundene Wechselwirkungsgeometrien, sondern eine den 3D-QSAR-Methoden nachempfundene Gittertechnik. Für jedes in den zu dockenden Liganden vorhandene Element muss ein Gitter mit Interaktionsenergien *vor* dem eigentlichen Docking berechnet werden. Ebenso muss – nur einmal – ein Gitter für elektrostatische Wechselwirkungen erzeugt werden. Diese Berechnung im Voraus bedeutet einen erheblichen Geschwindigkeitsvorteil gegenüber einer Energieevaluierung simultan zum Dockingprozess. Für nicht mit Gitterpunkten zusammenfallende Atomkoordinaten (was häufig der Fall ist) werden die Energiewerte zwischen den Gitterpunkten interpoliert. Von besonderem Vorteil ist die Flexibilität für Dockingexperimente mit anderen Scoringfunktionen, da man prinzipiell in jede geeignet skalierte Energiematrix docken kann. So wurden beispielsweise die Potentiale der Scoring-Funktion DrugScore verwendet [Sotriffer et al., 2002], adaptierte proteinspezifische AFMoC-Potentiale [Radestock et al., 2005] und auch ein Docking in mehrere Gitter simultan [Sotriffer & Dramburg, 2005; Zentgraf et al., 2006], die von unterschiedlichen Proteinkonformationen oder Isoformen eines Proteins abgeleitet wurden, ist beschrieben worden.

Zur Suche von Energieminima sind in AutoDock die Methoden *Monte Carlo Simulated Annealing* (MC-SA), *Genetischer Algorithmus* (GA), *Lamarck'scher Genetischer Algorithmus* (LGA) und *Evolutionary Programming* (EP) verfügbar. Für die meisten Anwendungen hat sich die LGA-Methode bewährt.

GOLD

Das Programm GOLD (*Genetic Optimisation for Ligand Docking*) [Jones et al., 1995, 1997; Verdonk et al., 2003] verwendet ebenfalls einen Genetischen Algorithmus zur Suche nach den günstigsten Orientierungen und Konformationen von Liganden in der Proteinbindungsstelle. Neben der ursprünglichen Scoring-Funktion, GOLD-Score, ist auch ChemScore in GOLD verfügbar sowie die Möglichkeit, eine benutzerdefinierte Scoring-Funktion zu verwenden. Das Programm ermöglicht es, zahlreiche Zusatzbedingungen, sog. *constraints*, zu definieren, die eine potentielle Dockinglösung erfüllen muss, so dass Vorwissen über das Target mit einbezogen werden kann.

Eine neuere Funktionalität soll die differenziertere Behandlung von Wassermolekülen während des Dockings ermöglichen [Verdonk et al., 2005]: Wassermoleküle können entweder als vorhanden, nicht vorhanden oder – in Abhängigkeit von der Balance zwischen einem entropischen Strafterm für gebundenes Wasser und einem Energieterm für Wechselwirkungen mit dem Protein – als variabel definiert werden. Im letzteren Fall entscheidet ein entsprechender Energieterm über den Verbleib eines Wassermoleküls. Ferner kann der Benutzer einstellen, ob ein Wassermolekül fixiert werden soll oder ob eine Rotation um seine Hauptachsen zur Optimierung des Wasserstoffbrückennetzwerkes erlaubt sein soll.

2.4.3 FlexS

Bei dem Programm FlexS [Lemmen & Lengauer, 1997; Lemmen et al., 1998] handelt es sich um ein von der Proteinstruktur unabhängiges Alignmentverfahren. Der **Referenzligand** dient als Templat und wird rigide gehalten; ein weiterer, flexibel behandelte **Testligand** wird auf diesen überlagert. Wie bei den anderen Produkten der „Flex-Gruppe“ arbeitet auch FlexS mit einem inkrementellen Aufbaualgorithmus. Im Gegensatz zu Docking-Prozeduren, bei denen eine Komplementarität des Liganden mit dem Rezeptor angestrebt wird, werden bei diesem Verfahren Interaktionsgeometrien benutzt, um mögliche Alignments zu finden, bei denen Referenz- und Testligand dieselben gerichteten Wechselwirkungen mit einem hypothetischen Rezeptor ausbilden können. Neben diesem Kriterium wird eine maximale Überlappung bezüglich verschiedener Eigenschaften (Hydrophobizität, Partialladung, Wasserstoffbrückenbindungseigenschaften) angestrebt, die durch Gaussfunktionen im Raum be-

schrieben werden. Ferner wird versucht, ein maximales gemeinsames van-der-Waals-Volumen beim Alignment zu erreichen.

Als Referenzligand kann man beispielsweise aus Kristallstrukturen extrahierte Moleküle, Dockinglösungen, gemäß Kap. 2.4.1 generierte Geometrien oder aus Konformationsanalysen erhaltene Strukturen benutzen. Damit ist die Methode in An- und Abwesenheit einer Rezeptorstruktur anwendbar. Ein weiterer Vorteil ist die Erweiterung FLEXS^C, die ein schnelles Alignment großer kombinatorischer Molekülbibliotheken erlaubt, wobei einmal platzierte Fragmente im Laufe des Alignments wiederverwendet werden. Falls konservierte Ankerfragmente allen zu alignenden Molekülen gemeinsam sind, kann die Positionierung dieses Fragmentes als sog. MAPREF definiert und damit fixiert werden. Dadurch kann die Konsistenz des Alignments, welche für 3D-QSAR-Studien besonders wichtig ist, gewährleistet werden. Der Hauptvorteil gegenüber Docking besteht in der kaum noch erforderlichen Intervention durch den Benutzer: Zwar erzeugt auch diese Methode pro Ligand mehrere Lösungen, die durch eine Scoring-Funktion bewertet werden, jedoch ist die als am besten bewertete Lösung hierbei erfahrungsgemäß auch die konsistenteste im Hinblick auf die Referenz, während bei Docking/Scoring häufig – trotz gleichen Grundgerüsts der Liganden – vollkommen unterschiedliche Bindemoden als beste Lösungen vorhergesagt werden. Neben der Verwendung zum Alignment für 3D-QSAR-Untersuchungen können die Überlagerungen von FlexS beispielsweise zur Pharmakophoraufklärung oder – in Verbindung mit den Ähnlichkeitsscores – zum virtuellen Screening [Evers et al., 2005] verwendet werden.

2.5 Ausschließliche Verwendung von Proteininformation: GRID/CPCA

Im folgenden Abschnitt wird ein etabliertes Verfahren vorgestellt, das es erlaubt, ohne Zuhilfenahme von Ligandinformation selektivitätsbestimmende Faktoren durch Analyse der dreidimensionalen Struktur der Bindetaschen qualitativ zu extrahieren. Dazu werden zunächst die Bindetaschen der studierten Rezeptoren mit Hilfe eines gitterbasierten Verfahrens abgetastet. Wie bei den 3D-QSAR-Methoden werden die Wechselwirkungsenergien von Sondenatomen, die auf einem regelmäßigen Gitter platziert werden, mit den Proteinatomen berechnet. Die ursprünglichen Studien verwendeten hierzu das im *drug design* bewährte Programm GRID [Goodford, 1985; Wade et al.,

1993; Wade & Goodford, 1993; Carosati et al., 2004; Cruciani, 2005], das über viele verschiedenartige (Multiatom-)Sonden verfügt. Es besitzt ein eigenes Kraftfeld, welches sterische, elektrostatische, entropische und auch durch Flexibilität der Seitenketten des Targets bedingte Effekte berücksichtigt. GRID wurde auch zu den in dieser Arbeit durchgeführten Studien verwendet. Alternativ ist es möglich, andere gitterbasierte Verfahren zu verwenden, wie z. B. DrugScore oder das aus Kristallstrukturen kleiner organischer Moleküle abgeleitete SuperStar [Verdonk et al., 1999, 2001; Boer et al., 2001].

Für jedes Target wird ein Zeilenvektor erzeugt, der die Energiewerte an den einzelnen Gitterpunkten enthält. Diese Zeilenvektoren werden dann in einer X-Matrix zusammengefasst, der aber – im Gegensatz zu den 3D-QSAR-Verfahren – **keine** Y-Matrix mit Affinitäten gegenübersteht. Daher werden zur weiteren Analyse auch keine überwachten (engl.: *supervised*), wie die PLS-Analyse, sondern unüberwachte (engl.: *unsupervised*) Methoden angewendet. Ein solches unüberwachtes Verfahren ist die **Hauptkomponentenanalyse** oder **PCA** (engl.: *Principal Component Analysis*) [Wold, 1987; Migliavacca, 2003].

Die PCA stellt ein Verfahren dar, Daten aus höherdimensionalen Räumen in niedrigerdimensionale Räume zu projizieren, und behält dabei die Nachbarschaftsbeziehungen der Objekte zueinander bei. Die Methode wird häufig zur explorativen Analyse multivariater Daten verwendet, da bei einer Projektion in den zwei- oder dreidimensionalen Raum Ähnlichkeiten der Objekte visuell erkannt werden können. Nach geeigneter Aufbereitung der Rohdaten (oft durch Autoskalierung) wird die X-Matrix in eine **Score**-Matrix T und eine **Loading**-Matrix P zerlegt, so dass gilt:

$$T = X \cdot P \quad (2.8)$$

Besitzt die X-Matrix r Zeilen (entspricht z. B. r Bindetaschen) und s Spalten (entspricht z. B. s Deskriptoren), so besitzt die Score-Matrix r Zeilen und die Loading-Matrix s Zeilen. Die in der Score-Matrix abgelegten Werte werden als *Scores* bezeichnet, sie werden so berechnet, dass sie zueinander orthogonal sind, also keinerlei Redundanz beinhalten. Sie sind Linearkombinationen der ursprünglichen x -Variablen, gewichtet mit den Loadings. Je höher die ursprünglichen Variablen korreliert (redundant) sind, desto weniger Hauptkomponenten (PCs) sind erforderlich, um die Daten der X-Matrix zu beschreiben. Wieviele Hauptkomponenten ausreichend sind, wird quantifiziert, in-

dem die kumulative Varianz, die durch eine bestimmte Anzahl an PCs beschrieben wird, berechnet wird. Insofern stellt die PCA eine Methode zur Daten- und Rauschreduktion dar.

Gemäß obiger Ausführungen sind die Scores eine Art „neue“, unkorrelierte Variablen, die die Lage der Objekte in einem niedriger dimensional Raum darstellen. Häufig werden daher die Scores der ersten beiden oder ersten drei Hauptkomponenten in einem sog. **Score-Plot** gegeneinander aufgetragen; dicht beieinander liegende Objekte sind dabei im Eigenschaftsraum der X-Matrix ähnlich zueinander, weit auseinander liegende unähnlich. Im **Loading-Plot** hingegen entspricht jeder Punkt einem Deskriptor der X-Matrix. Hochkorrelierte Deskriptoren, die eine ähnliche Information enthalten, liegen dicht beieinander, unkorrelierte mit unterschiedlicher Information weit weg voneinander. Liegt ein zu einem Deskriptor gehörender Punkt im Loading-Plot nahe am Ursprung des Koordinatensystems, so besitzt dieser Deskriptor eine geringe Bedeutung für das PCA-Modell, weit vom Ursprung entfernte Punkte hingegen kennzeichnen Deskriptoren mit hoher Bedeutung für das Modell.

Die ersten Studien dieser Art wurden mit einer gewöhnlichen PCA durchgeführt [Pastor & Cruciani, 1995; Matter & Schwab, 1999]. Dabei wurden jeweils nur zwei Targets studiert und die X-Matrix bei n Sonden mit $2n$ Zeilen gefüllt, während diese k Spalten für k Gitterpunkte enthielt. Dieses Vorgehen hatte mehrere Nachteile: Abhängig von der Natur der Sonde besitzen hydrophobe Wechselwirkungen (C3- oder DRY-Sonde) niedrigere numerische Werte als beispielweise elektrostatische Wechselwirkungen (O- oder COO-Sonde). Durch die Anordnung der Werte in der X-Matrix kann keine Korrektur durch geeignete Skalierung vorgenommen werden, so dass hydrophobe Wechselwirkungen deutlich unterbewertet würden. Ferner müssen bei dieser Variante die für die Selektivität wichtigen Regionen der Bindetasche aus dem Loading-Plot identifiziert werden, der hierbei für jeden Gitterpunkt einen Punkt enthält. Durch Rückprojektion in die Bindetasche erhält man zwar Regionen, die wichtig für die Selektivität sind, nicht aber nach individuellen Interaktionstypen (Sonden) getrennt.

Aufgrund dieser Unzulänglichkeiten wurde die Methode der sog. **Consensus Principal Component Analysis** oder **CPCA** eingeführt [Kastenholz et al., 2000]. Dazu werden die Daten in folgender Weise organisiert (Abb. 2.4): Besteht das Gitter aus k Punkten und werden n verschiedene Sonden verwendet, so erhält man für jede Bindetasche einen $n \times k$ Zeilenvektor. Diese Zeilenvektoren werden untereinander angeordnet und zu einer X-Matrix zusammengefasst.

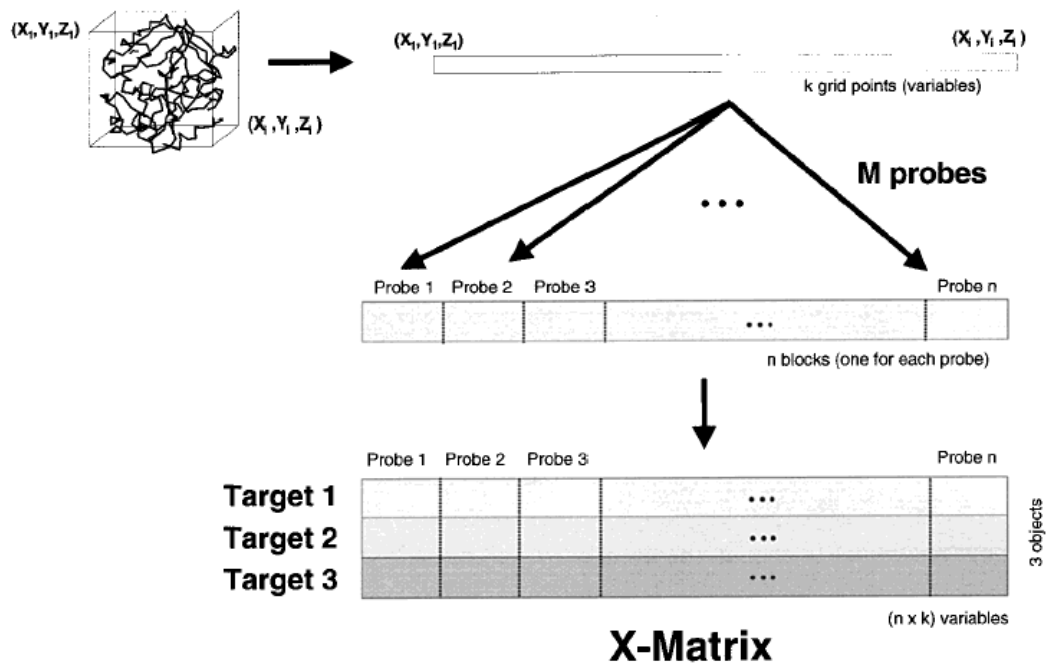


Abbildung 2.4: **Aufbau der X-Matrix zur Durchführung einer CPCA.** Ausgehend von einer GRID-Rechnung werden die Energiewerte, für k Gitterpunkte mit n Sonden berechnet, in einem $n \times k$ Zeilenvektor abgelegt. Jeder auf diese Weise für ein Target berechnete Zeilenvektor wird dann in der X-Matrix gespeichert (Abb. aus [Kastenholz et al., 2000]).

Der Hauptunterschied zur gewöhnlichen PCA besteht in der Organisation der Energiewerte in einzelnen **Blöcken** und darin, dass bei der CPCA nicht jedes Target-Interaktionstyp-Paar, sondern jedes einzelne Target ein Objekt darstellt. Diese Anordnung ermöglicht es, die Energiewerte blockweise, getrennt für jeden Interaktionstyp, zu skalieren (sog. *Block Unscaled Weighting, BUW*) und so allen Proben, unabhängig von ihrer numerischen Spannweite, initial denselben Einfluss auf das Modell zu geben. Ferner gibt es jeden von der PCA her bekannten Plot auf einem *Block-Level* (getrennt nach Sonden) und einem *Consensus-Level* (gleichzeitige Betrachtung aller Sonden). Weiterhin gibt es einen Plot, der die Gewichtung der einzelnen Blöcke (Sonden) für das Gesamtmodell beschreibt, den sog. *Superweights Plot* (die Terminologie wurde hier von dem Programm GOLPE übernommen, das üblicherweise für derartige Berechnungen verwendet wird). Diese Erweiterungen ermöglichen eine einfachere und differenziertere Identifizierung selektivitätsbestimmender Bereiche in den Bindetaschen und die der zugehörigen Interaktionstypen.

Alternativ zum GRID/CPCA-Ansatz wurden entsprechende Konturkarten erstellt, in-

dem die Energiewerte an entsprechenden Gitterpunkten voneinander subtrahiert und die Differenzwerte konturiert wurden. Die Ergebnisse wurden verglichen und diskutiert (siehe Kap. 4). Dabei wurde deutlich, dass die Konturkarten, welche durch einfache Differenzbildung erhalten wurden, nahezu identisch mit denen aus der GRID/CPCA-Methode sind.

2.6 Beispiele aus der Literatur

Der folgende Abschnitt soll einige Beispiele publizierter QSAR-Selektivitätsstudien vorstellen, kann aber keinesfalls eine vollständige Auflistung beinhalten. Eine Anwendung der klassischen Hansch-Analyse wird von Prasanna et al. beschrieben [Prasanna et al., 2004], worin sowohl getrennte Modelle für die Affinität zu COX-1 und COX-2 als auch für die Selektivitätsvariable (dem Quotienten für die inhibitorischen Konzentrationen COX-1/COX-2) erstellt werden. Bei den Liganden handelt es sich um annellierte Pyrazolderivate. López-Rodríguez et al. verwenden sowohl die klassische Hansch-Analyse als auch künstliche neuronale Netze, um jeweils separate Modelle für den 5-HT_{1A}- und den α_1 -adrenergen Rezeptor abzuleiten und damit Selektivitäten von Arylpiperazinen vorhersagen zu können [López-Rodríguez et al., 2001]. Ein Vergleich zwischen Hansch-Analyse und CoMFA-Modellen für die Vorhersage von Affinitäten zum 5-HT_{1A}- und zum 5-HT₂-Rezeptor von Tetrahydropyridinindolen wird von Agarwal et al. beschrieben [Agarwal et al., 1993].

Ein frühes Beispiel für die Verwendung von Affinitätsdifferenzen in 3D-QSAR-Studien zur Vorhersage von Selektivität ist die Untersuchung von Wong et al. Sie leiteten einerseits getrennte CoMFA-Modelle für Diazepam-empfindliche und -unempfindliche Benzodiazepin-Rezeptoren für eine Serie von Imidazodiazepinen ab. Andererseits wurde die Affinitätsdifferenz als abhängige Variable benutzt, um zu einem Selektivitätsmodell zu gelangen [Wong et al., 1993]. Diese Vorgehensweise der Verwendung von Affinitätsdifferenzen ist in der Literatur in zahlreichen Beispielen zu finden; im Folgenden einige Beispiele für Anwendungen aus der jüngeren Vergangenheit: Dessalew et al. führen separate und Differenzstudien an Bisarylmaleimid-Inhibitoren für Glykogen Synthase Kinase 3, CDK2 und CDK4 durch, wobei die Differenzmodelle den Einzelmodellen sogar als statistisch überlegen beschrieben werden [Dessalew & Bharatam, 2007]. Eine Selektivitätsstudie für die muskarinischen M₁- und M₂-Rezeptoren an einer Serie von Piperidinylpiperidinen wurde von Niu et al. durchgeführt [Niu et al., 2007].

Zur Entwicklung D₃-selektiver Radioliganden zur Positronen-Emissions-Tomographie (PET) wurden CoMFA- und CoMSIA-Modelle mit den paarweisen Affinitätsdifferenzen zwischen D₃/D₂- und D₃/D₄-Rezeptoren erstellt [Salama et al., 2007]. Die Modelle wurden erfolgreich zum Design selektiver, D₃-affiner PET-Liganden eingesetzt. Böhm et al. untersuchten die selektivitätsbestimmenden Eigenschaften von Benzamidinen für die Serinproteasen Thrombin, Trypsin und Faktor Xa [Böhm et al., 1999]. Dabei wurden abermals Einzel- und Differenzmodelle mit der CoMFA- und CoMSIA-Methode erstellt.

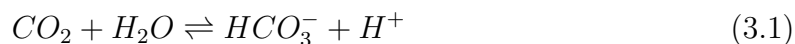
Die rein proteinbasierten Selektivitätsstudien wurden zunächst mit gewöhnlicher PCA, angewendet auf die GRID-Deskriptoren, durchgeführt. Die Targetenzyme waren humane/bakterielle Dihydrofolatreduktase [Pastor & Cruciani, 1995] und die Matrix Metalloproteinasen MMP-3 und MMP-8 [Matter & Schwab, 1999]. Im Falle der letzteren Studie wurden auch CoMFA-, CoMSIA- und GRID/GOLPE-Modelle jeweils einzeln und für den Quotienten der IC₅₀-Werte der Enzyme erstellt, so dass ein Vergleich zwischen den ligand- und proteinbasierten Methoden gezogen werden konnte. Mit der Implementierung der CPCA in das Programm GOLPE wurden proteinbasierte Studien meist mit dieser verbesserten Technik durchgeführt. Das Verfahren wurde erstmals an den Serinproteasen Trypsin, Thrombin und Faktor Xa demonstriert [Kastenholz et al., 2000]. Dieses System wurde eingehend von Böhm untersucht, wobei er neben GRID- auch SuperStar- und DrugScore-Deskriptoren verwendete und auch die Eignung von simpler Differenzbildung der Energiewerte – anstelle einer differenziellen CPCA – untersuchte [Böhm, 2002]. Naumann und Matter veröffentlichten eine ausführliche Selektivitätsstudie, in der die Interaktionsprofile verschiedener Kinasefamilien mit dem GRID/CPCA-Verfahren untersucht wurden und verglichen die erhaltenen Konturkarten mit parallel generierten CoMFA- und CoMSIA-Modellen, die von einer Serie von Purinderivaten für CDK1 abgeleitet wurden [Naumann & Matter, 2002]. Weitere GRID/CPCA-Studien wurden für Nitroxid-Synthase [Mattioni & Jurs, 2002], einen erweiterten Satz von Matrix Metalloproteinasen (MMP) [Pirard & Matter, 2006] (in dieser Studie wurden auch DrugScore *hot spots* zusätzlich zu GRID-Deskriptoren eingesetzt), Isoformen der nukleären *Peroxisome Proliferator-Activated Receptors* (PPARs) [Pirard, 2003] und Homologiemodelle bestimmter Cytochrom P450 2C-Isoformen durchgeführt [Ridderström et al., 2001].

3 Carboanhydrase als Modellsystem

3.1 Klassifizierung, Struktur und katalytischer Mechanismus

Carboanhydrasen (CAs, EC 4.2.1.1) sind ubiquitäre, zinkhaltige Enzyme, die in Pro- und Eukaryoten vorkommen. CAs werden von vier evolutionär nicht verwandten Genfamilien kodiert: α -CAs kommen in Wirbeltieren, Bakterien, Algen und dem Cytoplasma grüner Pflanzen vor [Smith & Ferry, 2000], β -CAs hauptsächlich in Bakterien, Algen und den Chloroplasten von *Mono*- und *Dicotyledoneae*, γ -CAs überwiegend in Archaeobakterien und einigen Eubakterien, während δ -CAs in einigen marinen Diatomeen zu finden sind [Supuran & Scozzafava, 2007].

Die katalytisch aktiven Formen der CAs katalysieren die reversible Hydratisierung von Kohlendioxid zu Hydrogencarbonat unter Freisetzung eines Protons [Maren, 1967; Lindskog, 1997]:



Daher sind CAs an einer Vielzahl bedeutsamer (patho)physiologischer Reaktionen beteiligt, z.B. Atmung und Transport von CO_2 /Bicarbonat zwischen metabolisierenden Geweben und Lunge, pH- und CO_2 -Homöostase, Elektrolytsekretion, biosynthetische Reaktionen (Glukoneogenese, Lipogenese, Ureagenese), Knochenresorption und -calcifizierung und Tumorigenizität [Supuran et al., 2004, 2003; Liljas et al., 1994, 1972; Lane & Morel, 2000; Hilvo et al., 2005].

Neben der Reaktion Gl. 3.1 katalysieren die α -CAs eine Reihe anderer Reaktionen, beispielsweise die Hydratisierung von Cyanat zu Carbamat, von Cyanamid zu Harnstoff, von Aldehyden zu geminalen Diolen und die Hydrolyse von Carbon- oder Sulfonsäureestern [Briganti et al., 1999; Guerri et al., 2000; Supuran et al., 1997]. Es ist bisher unklar, ob derartige Reaktionen eine physiologische Relevanz besitzen.

Die α -CAs besitzen stets eine monomere Einheit, β -CAs hingegen kommen als Dimer,

Tetramer, Hexamer oder Oktamer vor, und die γ -CA aus *Methanosarcina thermophila* bildet ein Trimer. Allen aktiven CAs ist das Vorhandensein eines Zinkions gemeinsam, das aber bei verschiedenen Familien von unterschiedlichen Aminosäuren komplexiert sein kann. Am katalytischen Mechanismus ist stets ein Wassermolekül beteiligt, das aber nicht in allen Fällen direkt an das Metall gebunden sein muss. Diese Details werden im folgenden Abschnitt für jede Familie getrennt diskutiert.

3.1.1 α -Carboanhydrasen

Die Klasse der α -Carboanhydrasen umfasst sämtliche humanen CAs; alle in dieser Arbeit durchgeführten Studien wurden ausschließlich an dieser Familie durchgeführt. In allen aktiven α -CAs ist Zn^{2+} als für die Katalyse essentielles Metallion enthalten. Dieses wird, am Boden einer konischen, 15 Å tiefen Bindetasche lokalisiert, von drei Histidinresten (His94, His96 und His119 gemäß CA I-Nummerierung) und einem Wassermolekül/Hydroxidion koordiniert. CAs der α -Klasse besitzen ein Molekulargewicht von ca. 30 kDa und bestehen typischerweise aus einem zehnsträngigen, antiparallelen β -Faltblatt und einigen helikalen Bereichen (Abb. 3.1).

Das zinkgebundene Wassermolekül bildet eine Wasserstoffbrücke zur Hydroxylgruppe des Thr199 aus, welche ihrerseits als Donor in einer Wasserstoffbrücke zum Glu106 fungiert (Abb. 3.2).

Der katalytische Mechanismus für die α -CAs wurde eingehend untersucht und stellt sich wie folgt dar:

In der aktiven Form des Enzyms liegt das Wassermolekül als Hydroxidion vor, welches an das Zinkion gebunden ist (Abb. 3.3, A, Seite 30). Im nächsten Schritt wird CO_2 in einer benachbarten hydrophoben Tasche angelagert, die von den Aminosäuren Val121, Val143 und Leu198 (bei hCA II) gebildet wird (Abb. 3.3, B). Dieses wird von dem Hydroxidion nukleophil angegriffen, so dass ein ans Zinkion gebundenes Hydrogencarbonation gebildet wird (Abb. 3.3, C). Das Bicarbonation wird dann durch ein Wassermolekül verdrängt, wodurch die saure Form des Enzyms gebildet wird (Abb. 3.3, D). Durch die Zinkkoordination wird der pK_S -Wert des Wassermoleküls von ca. 14 auf 7 herabgesetzt [Tripp et al., 2001], wodurch die nachfolgende Abspaltung eines Protons und damit die Regeneration des Hydroxidions ermöglicht wird.

Der geschwindigkeitsbestimmende Schritt bei dieser Reaktion ist der Protonentransfer

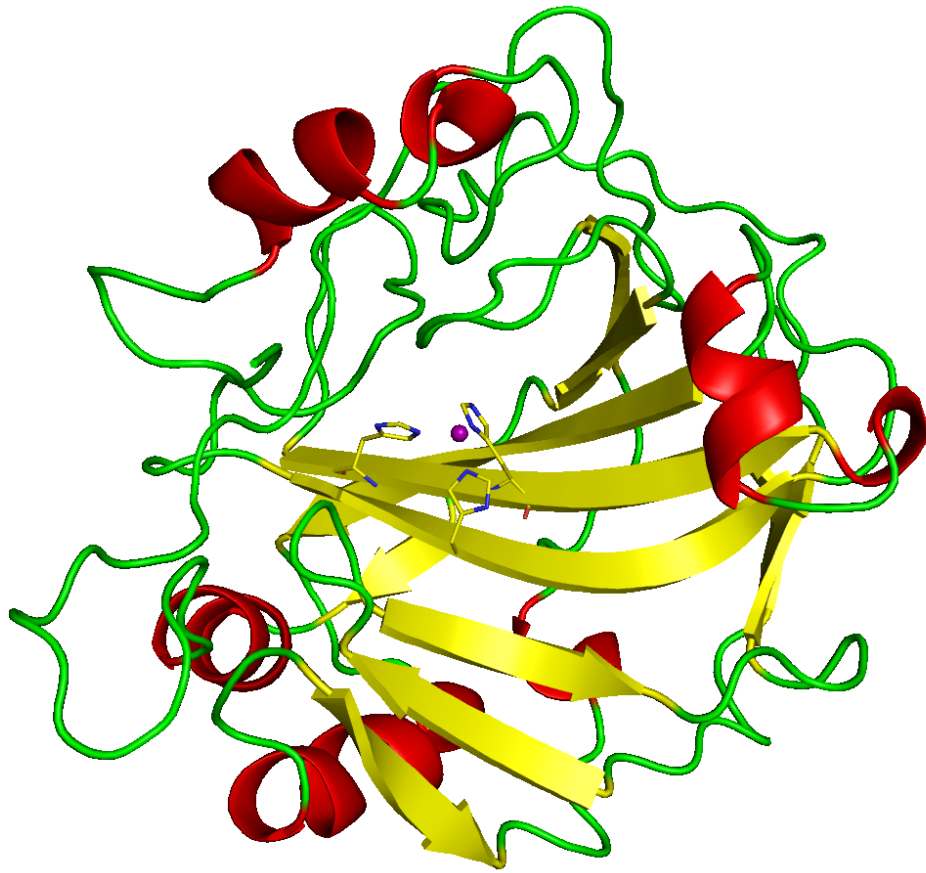


Abbildung 3.1: Struktur der α -Carboanhydrasen. Die Abb. zeigt als Beispiel die humane CA II (PDB-Code: 1CIL). In der Mitte ist das aktive Zentrum mit den drei Histidinresten, die das Zinkion (magenta) koordinieren, zu sehen. Typisches Sekundärstrukturelement ist das antiparallele, zehnsträngige β -Faltblatt (gelb). Helices sind in rot, Schleifenbereiche in grün dargestellt.

[Briganti et al., 1997]; bei den hochaktiven Isoformen, wie CA II, CA IV, CA VII und CA IX, wird dieser Prozess durch einen Histidin-Rest am Eingang des aktiven Zentrums unterstützt, His64, das deshalb auch als *proton shuttle* bezeichnet wird (Abb. 3.4, S. 31). Bei dem aktivsten der CA-Isoenzyme, CA II, wird die Effizienz des Protonentransfers durch einen Cluster von Histidinen erhöht, welcher sich vom Rand des aktiven Zentrums zur Oberfläche des Proteins erstreckt. Dies erklärt auch, warum CA II eines der schnellsten bekannten Enzyme überhaupt ist mit einer Wechselzahl von 10^6s^{-1} . Damit erreicht das Enzym die Grenze der Diffusionskontrolle [Christianson & Cox, 1999].

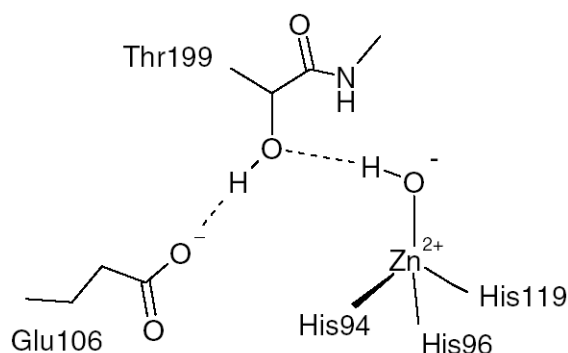


Abbildung 3.2: Koordination des Hydroxidions in der aktiven Form der hCA II.

Das Ion ist an $\text{Zn}(\text{II})$, welches von His94, His96 und His119 koordiniert wird, gebunden. Gleichzeitig bildet es eine Wasserstoffbrücke zur Hydroxylgruppe des Thr199 aus, die ihrerseits eine Wasserstoffbrücke zum Glu106 besitzt (Abb. aus [Supuran & Scozzafava, 2007]).

In den α -CAs führt ein Ersatz des Zn^{2+} -Ions durch Co^{2+} zu einer Halbierung der Aktivität, Substitution durch Co^{3+} , Cu^{2+} oder Ni^{2+} führt zum völligen Verlust der Enzymaktivität [Kumar & Kannan, 1994]. Wie oben bereits erwähnt, kommen α -CAs nicht nur in Wirbeltieren vor; so wurde durch die Kristallstrukturbestimmung einer CA aus *Neisseria gonorrhoeae* [Huang et al., 1998; Smith & Ferry, 2000] derselbe Aufbau wie in humanen α -CAs gefunden. Auch der Komplex mit dem Inhibitor Acetazolamid weist einen ähnlichen Bindungsmodus wie bei der hCA II auf. Ein weiteres Beispiel für eine prokaryotische α -CA wurde in *Helicobacter pylori* gefunden [Chirica et al., 2001]. Neben den katalytisch aktiven CAs wurden im menschlichen Organismus auch drei inaktive Formen entdeckt, die sog. *carbonic anhydrase related proteins* (**CARPs**). Dazu zählen CARP VIII, CARP X und CARP XI. Der Verlust der katalytischen Aktivität liegt darin begründet, dass den CARPs mindestens eines der drei zur Zinkbindung erforderlichen Histidine fehlt, so dass sie kein Zn^{2+} binden können. Durch Mutation ist es allerdings möglich, diese Proteine in aktive Enzyme umzuwandeln [Supuran & Scozzafava, 2007; Tashian et al., 2000].

Eine Übersicht über alle in höheren Wirbeltieren vorkommenden 16 α -CAs, ihre katalytische Aktivität, die Affinität zu Sulfonamidinhibitoren und ihre subzelluläre Lokalisation zeigt Tab. 3.1.

Tabelle 3.1: Übersicht über die α -CAs höherer Wirbeltiere, ihre relative CO₂-Hydratisierungsaktivität, die Affinität für Sulfonamidinhibitoren und die subzelluläre Lokalisation (nach [Supuran & Scozzafava, 2007]).

Isoenzym	Katalytische Aktivität	Affinität für Sulfonamide	subzelluläre Lokalisation
CA I	moderat	mittel	Zytosol
CA II	hoch	sehr hoch	Zytosol
CA III	sehr niedrig	sehr niedrig	Zytosol
CA IV	hoch	hoch	membrangebunden
CA VA	niedrig–moderat ^{a)}	hoch	Mitochondrien
CA VB	hoch	hoch	Mitochondrien
CA VI	moderat	hoch	Sekretion
CA VII	hoch	sehr hoch	Zytosol
CARP VIII	inaktiv	^{b)}	Zytosol
CA IX	moderat–hoch	hoch	transmembranär
CARP X	inaktiv	^{b)}	sezerniert
CARP XI	inaktiv	^{b)}	sezerniert
CA XII	niedrig	sehr hoch	transmembranär
CA XIII	moderat	mittel–hoch	Zytosol
CA XIV	moderat	hoch	transmembranär
CA XV	niedrig	unbekannt	membrangebunden

^{a)}niedrig bei pH 7.4, moderat bei pH 8.2 oder höher

^{b)}nicht messbar

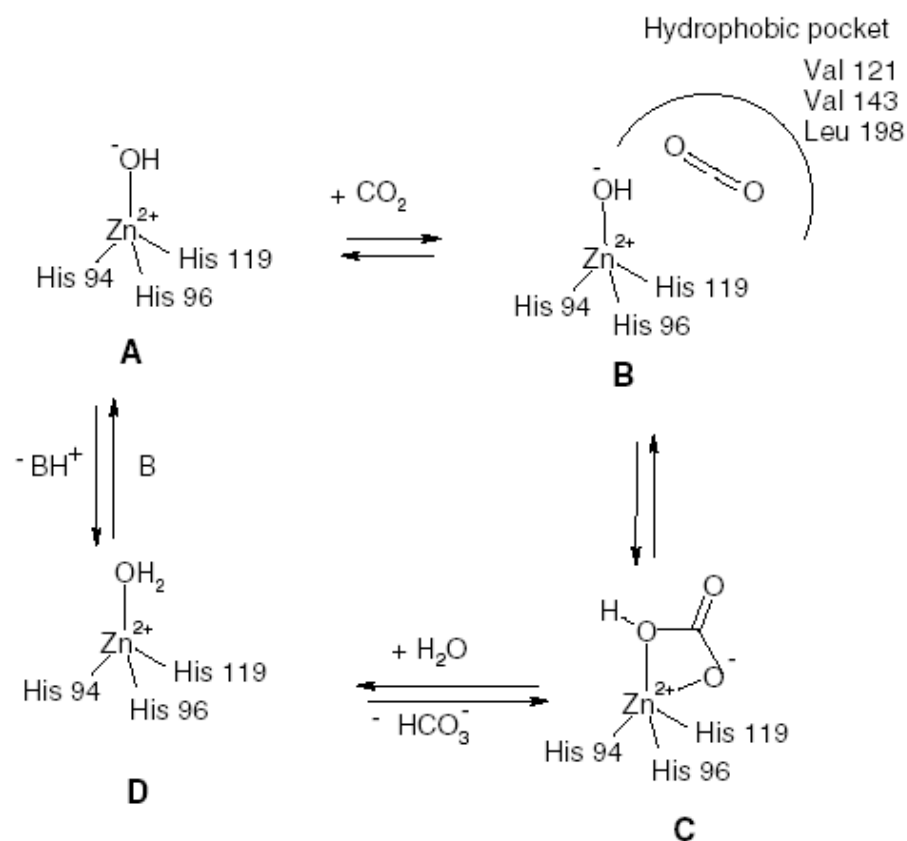


Abbildung 3.3: Katalytischer Mechanismus der CO_2 -Hydratisierung durch α -CAs.

In Schritt B ist schematisch die hydrophobe Tasche gezeigt, in der Substrate gebunden werden. (Abb. aus [Supuran & Scozzafava, 2007]).

3.1.2 β -Carboanhydrasen

Im Gegensatz zu den α -CAs sind die β -CAs üblicherweise als Oligomere organisiert, die von 2–8 Monomeren mit einem Molekulargewicht von 25–30 kDa gebildet werden. Aktuell (Stand: Juni 2007) sind vier Kristallstrukturen solcher β -CAs verfügbar: CAs aus der Rotalge *Porphyridium purpureum* [Mitsuhashi et al., 2000], aus den Chloroplasten von *Pisum sativum* [Kimber & Pai, 2000], aus *E. coli* [Cronk et al., 2001] und „cab“, isoliert aus dem Archaeobakterium *Methanobacterium thermoautotrophicum* [Smith & Ferry, 1999; Strop et al., 2001].

Bei den β -CAs wird das Zinkion nicht durch drei Histidine, sondern durch einen Histidin- und zwei Cysteinreste koordiniert. Als vierter Ligand tritt entweder ein Aspartatrest oder ein Wassermolekül/Hydroxidion auf: In der Kristallstruktur der β -CAs von *Porphyridium purpureum* und *E. coli* wurde ein Aspartatrest gefunden, während

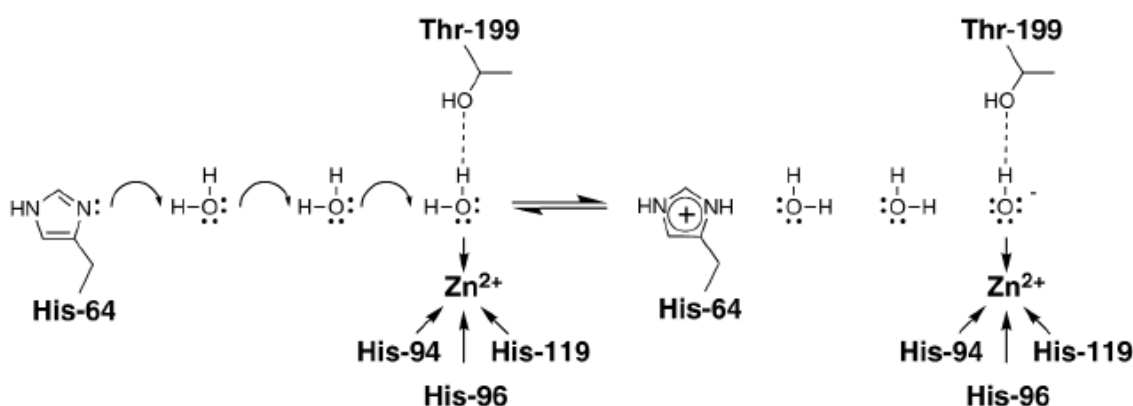


Abbildung 3.4: **Protonentransfer vom zinkgebundenen Wassermolekül zur Regeneration des Hydroxidions bei der CA II.** Die Ionisation findet durch Grotthuss-Diffusion eines Protons über weitere Wassermoleküle zum His64, dem *proton shuttle*, statt (Abb. aus [Christianson & Cox, 1999]).

bei *Pisum sativum* und *Methanobacterium thermoautotrophicum* ein Wassermolekül den vierten Liganden darstellt. Im ersten Fall muss der katalytische Mechanismus, wie er in Kap. 3.1.1 für die α -CAs diskutiert wurde, in veränderter Form stattfinden. Ein entsprechender modifizierter Mechanismus wurde von Mitsuhashi et al. [Mitsuhashi et al., 2000] vorgeschlagen und ist in Abb. 3.5 dargestellt.

Im Falle der β -CA von *Porphyridium purpureum* besteht ein Monomer aus zwei symmetrischen Motiven, von denen jedes ein Zinkion gebunden hat. Ein Wassermolekül ist hierbei nicht direkt an ein Zinkion koordiniert, sondern es bildet eine Wasserstoffbrücke zu dem Aspartatrest aus (Abb. 3.5, A). Es wird vermutet, dass ein Proton vom Wassermolekül auf das Aspartat übertragen wird, wodurch ein Hydroxidion gebildet wird. Dieses koordiniert zusammen mit der Carbonsäuregruppe des Asparaginsäurerestes an das Zinkion (Abb. 3.5, B), an dem eine bipyramidale Geometrie entsteht. Das dadurch gebildete, stark nukleophile Hydroxidion greift dann ein in der hydrophoben Tasche des Enzyms gebundenes Kohlendioxidmolekül an (Abb. 3.5, C), wodurch ein Hydrogencarbonation gebildet wird (Abb. 3.5, D). Dieses wird an die Lösung abgegeben, zusammen mit einem Proton (für den Transferprozess sind keine weiteren Details bekannt). Der dabei regenerierte Aspartatrest bindet erneut ein Wassermolekül, dadurch ist der Ausgangszustand wieder hergestellt, und das Enzym steht für einen neuen Katalyzyklus bereit.

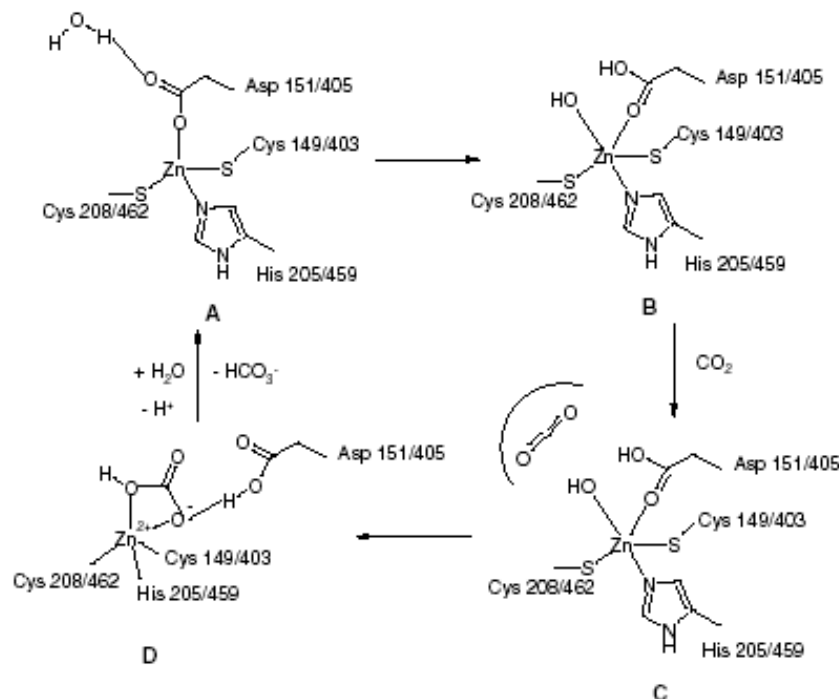


Abbildung 3.5: Postulierter katalytischer Mechanismus der CO_2 -Hydratisierung durch prokaryotische β -CAs (Aminosäurenummerierung gemäß *Porphyridium purpureum*) (Abb. aus [Supuran & Scozzafava, 2007]).

3.1.3 γ -Carboanhydrasen

Der Prototyp der γ -CAs, „Cam“, wurde aus dem Archaeobakterium *Methanosarcina thermophila* isoliert [Alber & Ferry, 1994]. Die Kristallstrukturen des zinkhaltigen und cobaltsubstituierten Enzyms sind aufgeklärt worden [Iverson et al., 2000]. Sie zeigen einige Unterschiede in der Faltung, verglichen mit den α - und β -CAs. Cam autoassoziiert zu einem Homotrimer mit einem ungefähren Molekulargewicht von 70 kDa. Das Zinkion wird von drei Histidinresten koordiniert, allerdings sind – im Ggs. zu den α -CAs – bei den γ -CAs weitere Wassermoleküle an das Metallion gebunden, so dass bei der zinkhaltigen Cam-Form eine trigonal-bipyramidale Geometrie und bei dem cobaltsubstituierten Enzym eine oktaedrische Anordnung resultiert. Zwei der metallkoordinierenden Histidine gehören zum einen Monomer, das dritte zu einem anderen, benachbarten. Daher sind die drei aktiven Zentren eines Trimers an den Kontaktflächen zwischen jeweils zwei monomeren Einheiten lokalisiert. Es wird angenommen, dass der katalytische Mechanismus der γ -CAs ähnlich zu dem der α -CAs (s. Kap. 3.1.1) verläuft; wahrscheinlich steht der pentakoordinierte Zustand am Zinkion in einem Gleichgewicht

mit einem tetraedrischen Zustand, wie er von den α -CAs her bekannt ist [Iverson et al., 2000].

3.1.4 δ -Carboanhydrasen und cadmiumhaltige CAs

In der marinen Diatomee *Thalassiosira weissflogii* wurde eine zinkhaltige CA mit einem Molekulargewicht von 27 kDa gefunden, die keine signifikanten Sequenzähnlichkeiten zu anderen CAs aufweist und daher einer neuen Klasse zugeordnet wird, den δ -CAs. Das Enzym wird als *Thalassiosira weissflogii* CA (TWCA1) bezeichnet; durch Röntgenabsorptionsspektroskopie an der Zn K-Kante konnte gezeigt werden, dass das Zinkion – wie bei den α -CAs – von drei Histidinresten und einem Wassermolekül koordiniert wird [Cox et al., 2000].

Interessanterweise wird bei Zinkmangel, erhöhtem Cadmiumangebot oder erniedrigtem CO₂-Partialdruck ein weiteres Protein mit Carboanhydraseaktivität mit einem Molekulargewicht von 69 kDa verstärkt exprimiert. Es ist das erste bekannte natürliche Enzym mit Cadmium im aktiven Zentrum und wird als CDCA1 bezeichnet. Röntgenabsorptionsspektroskopie zeigt, dass das Cadmiumion wahrscheinlich ungefähr in tetrahedraler Geometrie vorliegt und vermutlich – ähnlich den β -CAs – von zwei oder mehr Thiolaten und einem Wassermolekül koordiniert wird. CDCA1 besitzt keine signifikanten Sequenzähnlichkeiten mit anderen bekannten CAs und wurde in einer Publikation von Lane et al. [Lane et al., 2005] einer neuen Klasse, den ζ -CAs, zugeordnet. Ein solches cadmiumhaltiges Enzym übernimmt wahrscheinlich die für die Photosynthese wichtige CA-Funktion bei Mangel an Zinkionen (Meerwasser ist oft außerordentlich arm an Spurenelementen) [Lane & Morel, 2000; Lane et al., 2005].

3.2 Isoenzyme der humanen α -CAs

Der folgende Abschnitt gibt einen Überblick über die verschiedenen Isoenzyme der humanen α -CAs im Hinblick auf deren Lokalisation, die (patho)physiologische Funktion, die therapeutische Bedeutung einer Inhibition und ggf. strukturelle Besonderheiten.

3.2.1 CA I

Die CA I ist ein zytosolisches Enzym und im menschlichen Körper weit verbreitet. Die genaue Funktion dieses Enzyms ist bis heute unklar. Es scheint jedoch an Atmung, CO_2 -Transport und Säure-Base-Homöostase beteiligt zu sein [Maren, 1967; Supuran et al., 2003]. Die katalytische Aktivität der CA I ist etwa 100-fach niedriger als die der CA II. Mehrere Kristallstrukturen der CA I mit Sulfonamidinhibitoren [Chakravarty & Kannan, 1994] und ohne Liganden [Kannan et al., 1984] sind in der PDB hinterlegt.

Eine strukturelle Besonderheit der CA I ist ein Histidinrest in Position 200. Kein anderes Isoenzym weist an dieser Stelle ein Histidin auf, bei den meisten Isoformen ist dort ein Threonin vorhanden. Dies könnte für das Design selektiver CA I-Inhibitoren von Bedeutung sein. CA I wird von Anionen wie Cyanid, Thiocyanat, Cyanat oder Halogeniden stärker inhibiert als CA II, während die Affinität zu den Sulfonamidinhibitoren bei der CA I meist deutlich geringer ist als bei der CA II. Verbindungen, die eine höhere Affinität zu CA I aufweisen als zu CA II, gehören zur Gruppe der Phenylsulfonamide oder der heterozyklischen Sulfonamide und besitzen meist eine (Thio)Ureido-Gruppe [Supuran et al., 2004].

3.2.2 CA II

CA II ist ebenfalls ein zytosolisches Enzym. Für dieses Isoenzym wurden die meisten Kristallstrukturen mit verschiedenen Inhibitoren und verschiedenen Mutationen in der PDB deponiert. Es besitzt eine sehr hohe Affinität für Sulfonamidinhibitoren und die höchste katalytische Aktivität unter allen CA-Isoenzymen. CA II ist ubiquitär im Körper vorhanden und an zahlreichen physiologischen Prozessen beteiligt; die folgende Auflistung beinhaltet daher nur einige besonders hervorzuhebende Funktionen.

In sämtlichen Zellen, insbesondere aber in Erythrozyten, reguliert sie – neben CA I und IV – den Transport von CO_2 im Blut und zwischen den metabolisierenden Geweben durch die Umwandlung in Bicarbonat. Die CA II beeinflusst in den Nieren die Ausscheidung von H^+ und die Resorption von HCO_3^- und ist damit an der pH-Homöostase beteiligt. CA II ist in den Ziliarfortsätzen der Uvea des Auges lokalisiert und dort maßgeblich an der Sekretion von Bicarbonat ins Kammerwasser beteiligt. Daher ist das Isoenzym von Bedeutung für die Behandlung des Glaukoms [Supuran & Scozzafava, 2000b]. Weiterhin wird es in der Glia, nicht aber in Neuronen selbst in größerem

Ausmaß exprimiert. Dort ist es im *Plexus choroideus* an der Produktion der cerebrospinalen Flüssigkeit beteiligt. CA II kommt auch in großen Mengen in den Osteoklasten des Knochens vor; dort ist sie wichtig zur Bereitstellung von H^+ für die Knochenresorption. Eine Inhibition könnte daher eine Therapieoption bei Osteoporose darstellen. In den Belegzellen der Magenschleimhaut stellt CA II H^+ -Ionen zur Verfügung, die mit Hilfe einer ATP-getriebenen Protonenpumpe ins Magenlumen sezerniert werden.

Neben den vielen Therapieoptionen, die eine Inhibition der CA II bietet, stellt diese Isoform ein besonderes Problem beim Design klinisch anwendbarer CA-Inhibitoren dar: Durch die extrem breite Verteilung über verschiedene Gewebe und die Tatsache, dass die meisten bekannten CA-Inhibitoren auch CA II erheblich oder sogar stärker hemmen als andere Isoenzyme, ist es schwierig, die Nebenwirkungen von CA-Inhibitoren von der erwünschten Wirkung zu trennen [Supuran & Scozzafava, 2007].

3.2.3 CA III

Die genaue physiologische Funktion der zytosolischen CA III ist bislang nicht bekannt. Neben der Kristallstruktur des Ratten-Enzyms [Mallis et al., 2000] ist auch die der humanen CA III in der PDB hinterlegt [Duda et al., 2005; Elder et al., 2007]. CA III wird im Skelettmuskel und in den Adipozyten exprimiert [Spicer et al., 1990; Sly & Hu, 1995]. Es wird vermutet, dass sie am CO_2 -Transport im Skelettmuskel und der CO_2 -Diffusion zu den Gewebekapillaren beteiligt ist [Duda et al., 2002]. Ihre katalytische Aktivität ist relativ gering (etwa 0,3% der von CA II), da sie in Position 64 statt eines Histidins ein Lysin besitzt, das ein wesentlich schlechteres *proton shuttle* darstellt. In Position 198 weist CA III ein Phenylalanin auf, dadurch wird die Bindetasche extrem verengt, was eine sehr niedrige Bindungsaffinität des Isoenzyms für Sulfonamidinhibitoren zur Folge hat. Demzufolge stellt CA III kein Problem in Bezug auf die Selektivitätsentwicklung von Inhibitoren dar.

3.2.4 CA IV

Bei der humanen CA IV handelt es sich um ein membrangebundenes Enzym; es ist über einen Glykosylphosphatidylinositolrest (GPI), der an Ser284 gebunden ist, in den Membranen von Endothelzellen verschiedener Organe, z. B. in Lungenkapillaren, oder in apikalen Membranen von Nierenepithelzellen verankert [Lindskog, 1997]. CA IV besitzt

eine vergleichsweise hohe CO_2 -Hydratisierungsaktivität und Affinität zu Sulfonamidinhibitoren [Baird et al., 1997]. Sie übernimmt zahlreiche physiologische Aufgaben in Verbindung mit der zytosolischen CA II. Die Lokalisation an Membranen ist bedingt durch ihre Aufgabe, die $\text{CO}_2/\text{HCO}_3^-$ -Umwandlung zu beschleunigen und so den Transport von CO_2 durch Membranen zu ermöglichen. Daher ist die CA IV wichtig für Atmung, CO_2 -Transport, Rückresorption von Bicarbonat in der Niere und Bildung des Kammerwassers im Auge [Supuran et al., 2004].

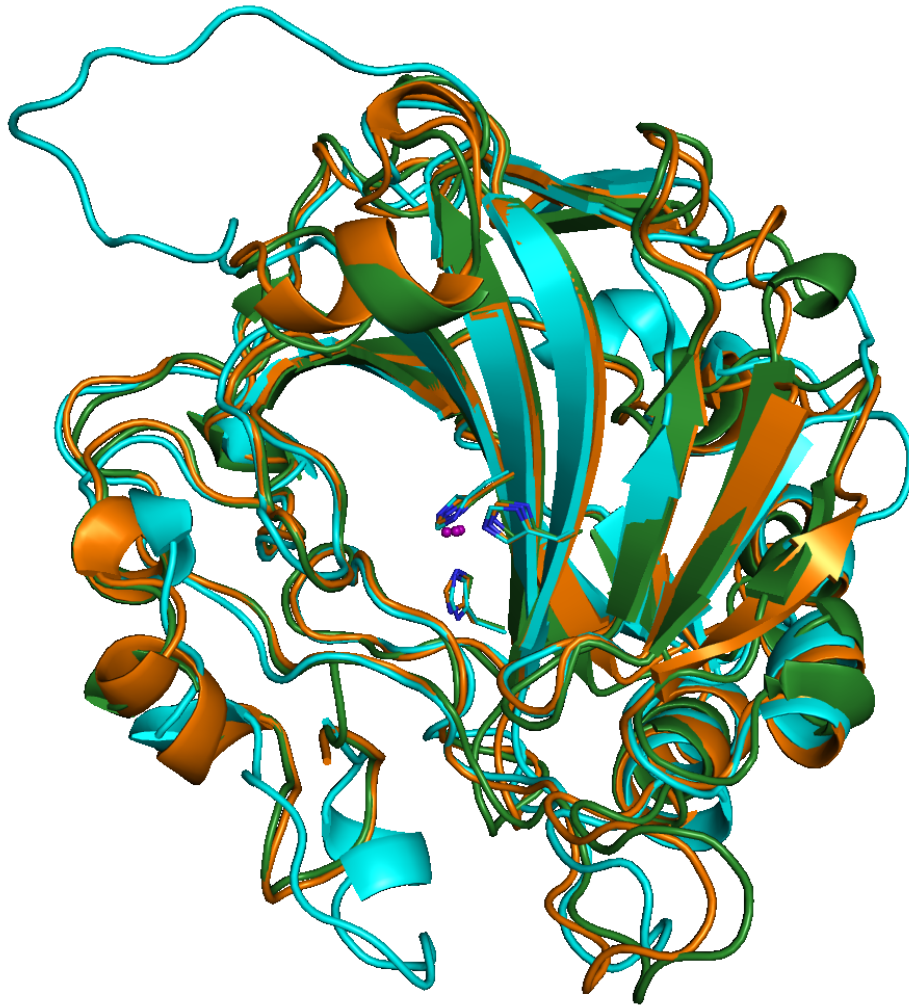


Abbildung 3.6: Überlagerung der drei humanen CA-Isoenzyme I (grün), II (orange) und IV (cyan). Im Zentrum sind die drei Histidinreste, die das katalytische Zinkion (violette Kugeln) koordinieren, hervorgehoben. Im linken oberen Bildbereich ist der im Text beschriebene Loop der CA IV deutlich erkennbar.

Die Kristallstruktur von löslicher CA IV, bei der der Glykosylphosphatidylinositolan-ker entfernt wurde, ist aufgeklärt worden. Als strukturelle Besonderheiten weist CA IV

zwei Disulfidbrücken (Cys6–Cys11 und Cys23–Cys203) in der Nähe des Bindetascheneingangs auf sowie einen flexiblen Loop (Lys124–Glu138), welcher solvensexponiert ist. In der entsprechenden Region in CA II sind die Aminosäuren Val135, Phe131 und Pro202 und interagieren meist mit hydrophoben Bereichen von Inhibitoren. Abb. 3.6 zeigt zum Vergleich die Überlagerung der drei humanen Isoenzyme CA I, II und IV. Aufgrund der abweichenden Orientierung des Loops von der Bindetasche weg entfallen diese Interaktionsmöglichkeiten, und die Tasche ist in dieser Region bei der CA IV deutlich weiter geöffnet.

3.2.5 CA VA und VB

Die Carboanhydrasen CA VA und VB sind die einzigen mitochondrialen CAs. Der Subtyp CA VA kommt beim Menschen in den Mitochondriden der Leber vor [Nagao et al., 1993], während VB in der Bauchspeicheldrüse und im Rückenmark nachgewiesen worden ist [Fujikawa-Adachi et al., 1999b]. Es wurde gezeigt, dass beide Isoformen an verschiedenen biosynthetischen Reaktionen beteiligt sind wie der Harnstoffsynthese [Dodgson, 1987], der Glukoneogenese [Dodgson & Cherian, 1989] und der Lipogenese [Lynch et al., 1995]. Sie erfüllen diese Aufgabe, indem sie für CO₂-abhängige Prozesse das Kohlendioxid in Form von Bicarbonat bereitstellen. An solchen Prozessen sind carboxylierende Enzyme beteiligt, wie Pyruvatcarboxylase, Acetyl-CoA-Carboxylase und die Carbamoylphosphatsynthetasen I und II. Aufgrund dieser Bedeutung für den Fettsäurestoffwechsel erscheint eine Inhibition der Isoenzyme VA und VB als möglicher Therapieansatz zur Behandlung der Adipositas [Simone & Supuran, 2007]. Unterstützt wird diese Vermutung durch die Beobachtung, dass Topiramat und Zonisamid, Antiepileptika mit potenter CA V-Inhibition, bei den Patienten zu signifikantem Gewichtsverlust führen.

Die CA V-Isoenzyme besitzen eine mäßige katalytische Aktivität, werden aber stark von Sulfonamidinhibitoren gehemmt. Bisher wurden nur die Kristallstrukturen der CA V von der Maus aufgeklärt [Boriack-Sjodin et al., 1995; Heck et al., 1996; Jude et al., 2002], allerdings sind – basierend auf diesen – Homologiemodelle der humanen CA VA konstruiert worden [Vitale et al., 2007]. Die auffälligsten Unterschiede zur CA II sind in der Loop/ α -Helix/Loop-Struktur im Bereich der Aminosäuren Asn124–Gly140 zu finden. Eine Insertion von Ser125 verursacht eine Verlängerung des Loops Asn124–Glu129, während die Mutationen Phe131Tyr und Cys206Ser möglicherweise das Was-

serstoffbrückenmuster der mutierten Reste so verändern, dass sich die Orientierung der Loop/Helix/Loop-Region ändert. Dadurch kommt es zu einer Verengung der Bindetasche von CA VA im Vergleich zu CA II, die noch weiter durch die Mutation Ala65Leu verstärkt wird.

3.2.6 CA VI

Die CA VI ist die einzige sezernierte Isoform, sie wird in Speichel, Tränenflüssigkeit und Muttermilch ausgeschieden. Bisher ist keine Kristallstruktur dieses Enzyms veröffentlicht worden. Die physiologische Bedeutung dieser CA ist bislang rätselhaft; es wird angenommen, dass sie eine Rolle bei der Geruchs- und Geschmackswahrnehmung sowie der pH-Regulation im Mundraum spielt. Es wird vermutet, dass der bittere, metallische Nachgeschmack oral angewandter CA-Inhibitoren auf eine Hemmung der im Speichel vorhandenen CA VI zurückzuführen ist [Supuran & Scozzafava, 2000b; Nishimori et al., 2007]. Dies könnte auch eine Erklärung für den metallischen Geschmack des Süßstoffes Saccharin sein, welcher ein submikromolarer Inhibitor von CA VI ist [Köhler et al., 2007]. Es wird kontrovers diskutiert, ob das Vorkommen von CA VI in Speichel und Zahnschmelz durch Beschleunigung der Dehydratisierung von Kohlensäure zur Neutralisation von durch Bakterien gebildeten Säuren beiträgt und CA VI damit ein „Anti-Kariesenzym“ darstellt [Kimoto et al., 2006] oder ob sie vielmehr durch Bildung von Protonen bei der CO₂-Hydratisierung den pH-Wert im Mundraum erniedrigt, so dass ihre Inhibition zur Kariesprophylaxe ausgenutzt werden könnte [Nishimori et al., 2007].

Mit einer katalytischen Aktivität, die etwa 33% derjenigen von CA II entspricht, ist ihre Geschwindigkeit moderat und mit der von CA I und IX vergleichbar. Möglicherweise kann die verringerte Katalysegeschwindigkeit durch eine Ala65Thr-Mutation erklärt werden; im Vergleich zur CA II behindert der sterisch anspruchsvolle Threoninrest die Flexibilität des Proton-Shuttles His64 und verzögert damit den schnellen Abtransport von Protonen. CA VI wird von Sulfonamidinhibitoren relativ stark gehemmt.

3.2.7 CA VII

Die Carboanhydrase VII ist ein zytosolisches Enzym und besitzt die zweithöchste katalytische Aktivität dieser Familie nach CA II (die Aktivität der CA VII beträgt ca.

70% der CA II-Aktivität) [Vullo et al., 2005]. Das Enzym wird in geringem Ausmaß in den verschiedensten Geweben exprimiert [Earnhardt et al., 1998], in besonderem Ausmaß jedoch im Gehirn. Eine Kristallstruktur der CA VII ist bislang nicht aufgeklärt worden. Die genaue Funktion der CA VII ist nicht gesichert, allerdings wird vermutet, dass sie wichtig für die Bereitstellung von Bicarbonat im Zytosol von Neuronen ist. Es wurde gezeigt, dass die CA VII-Expression im Hippocampus von Ratten sich in den ersten postnatalen Wochen drastisch verändert; bei erhöhter CA VII-Aktivität in Verbindung mit hochfrequenter Stimulation von pyramidalen Neuronen kann es zu einer Umkehr der sonst inhibitorischen Qualität der GABAergen Transmission kommen, so dass eine exzitatorische Wirkung durch GABA-Stimulation resultiert. Diese Vorgänge sind gekoppelt an komplizierte Verschiebungen der Anionenkonzentrationen und des pH-Wertes im intra- und extrazellulären Raum. Im Verlauf dieser Prozesse verlassen die durch CA VII freigesetzten HCO_3^- -Ionen die Pyramidenzellen durch den GABA_A -Rezeptor (ein ionotroper Neurorezeptor, der nicht nur für Cl^- , sondern auch HCO_3^- durchlässig ist). Dadurch wird auch der extrazelluläre pH-Wert beeinflusst [Thiry et al., 2007] .

Diese Erkenntnisse können wenigstens teilweise die Wirksamkeit einiger klinisch angewendeter CA-Inhibitoren bei speziellen Formen von Epilepsie erklären. CA VII wird von sulfonamidartigen Verbindungen sehr stark gehemmt. Neben CA VII sind aber wahrscheinlich auch die Isoenzyme CA II, IV, V, XII und XIV an neuronalen Prozessen beteiligt [Thiry et al., 2007].

3.2.8 CA IX und XII

Die CAs IX und XII sind transmembranäre Proteine; ihre katalytisch aktiven CA-Domänen sind extrazellulär lokalisiert; darauf folgt ein Transmembransegment und ein kurzer, intrazellulärer C-Terminus. CA IX weist am extrazellulären N-Terminus als Besonderheit eine sog. *proteoglycanähnliche* (PG) Region auf, die ihr eine Funktion zur Zelladhäsion verleiht [Závada et al., 2000; Svastová et al., 2003]. Die katalytischen Domänen von CA IX und XII, exprimiert in Bakterienzellen, besitzen eine CO_2 -Hydratisierungsaktivität, die 55% bzw. 23% derjenigen von CA II beträgt [Pastorekova et al., 2007]. CA IX wird bei normalen Geweben hauptsächlich im Verdauungstrakt exprimiert [Pastorekova et al., 1997], d.h. in Epithelzellen des Magens, des Dünndarms, des Dickdarms, der Speiseröhre, des Pankreas, der Gallenblase und des Gallengangs

[Supuran et al., 2004]. Dort ist sie stets in den basolateralen Membranen verankert, was auf ihre Aufgaben für die interzelluläre Kommunikation, die Aufrechterhaltung der Zellintegrität und den basolateralen Ionentransport hindeutet. Ferner ist sie beteiligt an Morphogenese und Kontrolle der Zellproliferation und -differenzierung von Zellen der Magenschleimhaut. CA XII dagegen zeigt eine wesentlich breitere Verteilung in gesunden Geweben; sie wird in reproduktiven Organen, dem Dickdarm, den Nieren und dem Auge exprimiert. Das Isoenzym liegt stets als Dimer vor, besitzt allerdings nicht die PG-ähnliche Region der CA IX.

Beide Isoenzyme werden verstärkt in vielen Tumorgeweben exprimiert, oft auch nebeneinander. Die Transkription des *CA9*-Gens wird durch Hypoxie induziert. Dies geschieht über HIF-1, den *hypoxia inducible factor*, der an HRE (*hypoxia response element*) bindet. Dies ist eine Sequenz im Promotor des *CA9*-Gens, die die Transkription reguliert. Ein solches HRE befindet sich nicht im *CA12*-Promotor. Besonders CA IX wird als interessantes Target für Diagnose und Therapie von Tumoren angesehen [Pastorekova et al., 2007]. Die Überexpression von CA IX bei Tumoren bewirkt eine Erniedrigung des pH-Wertes in der Umgebung von Tumorzellen. Diese verstärkt die Invasivität der Tumoren und kann auch die Aufnahme von Zytostatika beeinträchtigen. Für CA XII wurde weiterhin gefunden, dass sie – neben CA II und IV – an der Erhöhung des Augeninnendruckes beteiligt ist, so dass ihre Inhibition auch eine Rolle für die Glaukomtherapie spielen dürfte [Wilkinson et al., 2007]. Sowohl CA IX als auch XII werden von Sulfonamidinhibitoren stark gehemmt. Durch ihre extrazelluläre Lokalisation stellen v.a. membranimpermeable Inhibitoren eine Möglichkeit dar, Selektivität im Hinblick auf eine mögliche Tumorthherapie zu erhalten. Bisher ist nur die Kristallstruktur der extrazellulären Domäne der humanen CA XII aufgeklärt worden [Whittington et al., 2001], von der CA IX existiert derzeit keine.

3.2.9 CA XIII, XIV und XV

Die Isoenzyme XIII, XIV und XV wurden in jüngster Zeit entdeckt und charakterisiert. CA XIII ist ein zytosolisches Enzym, während CA XIV und XV an Membranen gebunden sind.

CA XIII ist phylogenetisch am nächsten mit CA I, II und III verwandt. Sie wird besonders im menschlichen Gastrointestinaltrakt, der Niere und den Fortpflanzungsorganen exprimiert. Es wird vermutet, dass das Isoenzym durch Produktion von Bicarbonat die

Motilität der Spermien begünstigt. Auch im weiblichen Genitaltrakt, im Cervix und im Uterus, ist das Enzym aktiv und könnte für die Einstellung eines optimalen pH-Wertes sorgen, der zur Befruchtung der Eizelle erforderlich ist [Hilvo et al., 2007]. Derzeit ist keine Kristallstruktur des Proteins in der PDB hinterlegt. Kinetische Studien haben gezeigt, dass die katalytische Aktivität der CA XIII moderat ist und der von CA I und V gleicht. CA XIII wird von den üblichen Sulfonamidinhibitoren relativ stark gehemmt (K_i -Werte vergleichbar mit CA II). Sulfanilylsulfonamide haben sich als Inhibitor-Klasse mit CA XIII-Präferenz erwiesen [Lehtonen et al., 2004].

CA XIV besitzt ein transzelluläres Segment, welches das Enzym an der Membran verankert. Sie wird in verschiedenen Geweben wie Niere, Herz, Gehirn, Skelettmuskel, Leber und Lunge exprimiert [Fujikawa-Adachi et al., 1999a]. In der Niere ist das Enzym – neben anderen CAs – wahrscheinlich an der Reabsorption von Bicarbonat beteiligt. Die Kristallstruktur der extrazellulären murinen CA XIV in der Apoform und im Komplex mit Acetazolamid ist aufgeklärt worden [Whittington et al., 2004]. Es handelt sich um ein N-glykosyliertes Protein mit einer Disulfidbrücke. Die katalytische Domäne ähnelt stark der von CA XII, mit der die CA XIV am nächsten verwandt ist. Die murine CA XIV ist extrem salzresistent. Möglicherweise erklärt dies das gleichzeitige Auftreten von CA IV und XIV in den renalen proximalen Tubuli; der Sinn besteht nicht in funktioneller Redundanz, sondern in einer Anpassung an unterschiedliche Salzkonzentrationen. Die katalytische Aktivität der humanen CA XIV wurde als mäßig hoch bestimmt und ist der von CA XII vergleichbar [Nishimori et al., 2005a]. Zwar wird CA XIV durch einige Gruppen von Sulfonamidinhibitoren stark gehemmt, jedoch wurden bisher keine Verbindungen mit ausgeprägter CA XIV-Präferenz gefunden [Hilvo et al., 2007].

CA XV ist das jüngste beschriebene CA-Isoenzym, es wurde 2005 entdeckt [Hilvo et al., 2005]. Es ist ein durch GPI-Anker an Membranen gebundenes Protein, wie die CA IV, der es auch phylogenetisch am nächsten steht. Eine Besonderheit der CA XV besteht darin, dass sie in zahlreichen Spezies exprimiert wird, nicht jedoch in Schimpansen und Menschen. Dort ist sie nur als Pseudogen vorhanden, welches nicht exprimiert wird. Bisher ist nur das Expressionsmuster in Mausgeweben charakterisiert worden: Dort ist sie in Niere, Gehirn, Hoden und 7–15 Tage alten Embryos zu finden. Die murine CA XV besitzt eine geringe katalytische Aktivität. Vermutlich steht das Enzym in funktioneller Redundanz mit CA IV und XIV [Hilvo et al., 2007].

3.2.10 CA-verwandte Proteine (CARP)

Außer den katalytisch aktiven CAs gibt es auch Proteine mit CA-ähnlicher Sequenz, aber ohne katalytische Aktivität. Diese werden als *carbonic anhydrase related proteins* (CARPs) bezeichnet [Tashian et al., 2000]. Der Nummerierung in der Reihenfolge ihrer Entdeckung folgend, werden die drei Isoformen als CARP VIII, X und XI bezeichnet. Den CARPs fehlen ein oder mehrere Histidinreste, die zur Zinkbindung erforderlich sind. Daher sind sie katalytisch inaktiv im Sinne der klassischen CO₂-Hydratisierung. Allerdings konnte durch entsprechende Mutationen die Aktivität der Maus-CA VIII wiederhergestellt werden [Elleby et al., 2000]. Ferner enthalten zwei Mitglieder der Familie der Protein-Tyrosinphosphatasen (PTPs), die *receptor-type protein tyrosine phosphatase* β (RPTP β = PTP ζ) und RPTP γ , im N-Terminus eine CA-ähnliche Domäne [Barnea et al., 1993; Levy et al., 1993]. Aufgrund des Fehlens zweier zinkbindender Histidine besitzen auch sie keine CA-Aktivität. Die biologische Funktion der CARPs ist bisher unbekannt.

Die CARP VIII, X und XI sind wahrscheinlich zytosolische Proteine. Sie kommen in erster Linie im Gehirn vor; CARP VIII wurde darüberhinaus auch in Herz, Plazenta, Skelettmuskel, Niere und Pankreas nachgewiesen. CARP IX ist außer im Gehirn auch noch in der Niere, CARP XI nur im Gehirn nachgewiesen worden [Supuran et al., 2004]. Es wurde beobachtet, dass CARP VIII in bestimmten Tumoren überexprimiert wird, bei nichtkleinzelligem Lungenkrebs und Kolorektalkrebs [Akisawa et al., 2003]. Es wird angenommen, dass CARPs Protein-Protein-Wechselwirkungen eingehen [Supuran et al., 2004]. Bislang sind keine Kristallstrukturen von CARPs aufgeklärt worden.

3.3 Carboanhydraseinhibitoren

Dieser Abschnitt soll einen kurzen Überblick über die verschiedenen Klassen von Carboanhydraseinhibitoren (CAI), ihrem Bindungsmodus und dem therapeutischen Einsatz zugelassener Verbindungen geben. Alle bekannten CA-Inhibitoren besitzen eine Funktionalität, mit der sie an das katalytische Zinkion koordinieren.

3.3.1 Substanzklassen

Die chemisch gesehen primitivsten CA-Inhibitoren sind bestimmte **Anionen** (z. B. OH^- , SH^- , Halogenide, CN^- , CNO^- , SCN^- , N_3^- , HSO_3^- , HCO_3^- , zahlreiche Anionen organischer Säuren). Sie sind sehr schwache Inhibitoren (im mM- oder μM -Bereich) und binden stets an das Zinkion. Je nach Anion gibt es drei verschiedene Möglichkeiten, wie die Koordination an das Zinkion stattfinden kann: Das aktivierte Wassermolekül kann verdrängt werden, so dass eine tetraedrische Geometrie entsteht, das Anion kann als fünfter Ligand unter Ausbildung einer verzerrten trigonal-bipyramidalen Geometrie an das Zn^{2+} -Ion koordinieren, oder es stellt sich ein Gleichgewicht zwischen beiden Geometrien ein. So verdrängen CN^- , CNO^- , SH^- , und Liganden wie Anilin, Anthranilat, Trichloroacetaldehyd, Thiadiazol, 1,2,4-Triazol, Tetrazol, Imidazol (bei hohem pH) und alle Sulfonamide das Nukleophil OH^- unter Ausbildung einer tetraedrischen Struktur. Im Gegensatz dazu führen Anionen wie SCN^- , HSO_3^- , NO_3^- , I^- , Formiat, Acetat, Oxalat, Malonat, Succinat, Glutarat und Liganden wie Glycin, L(+)-Alanin, D(-)-Alanin, 2,4-Pentandion und 1,2,3-Triazol zur Ausbildung pentakoordinierter, trigonal-bipyramidaler Spezies. Zwischen diesen beiden Extremen sind HCO_3^- , F^- , Cl^- , Br^- , N_3^- , Phosphat, Benzoat und Imidazol (bei niedrigem pH) anzusiedeln; sie bilden eine Mischung aus tetra- und pentakoordinierten Spezies. In Tab. 3.2 sind die Inhibitionskonstanten einiger Anionen für drei CA-Isoenzyme, bezogen auf die CO_2 -Hydratisierungsreaktion, gezeigt.

Einen recht außergewöhnlichen Inhibitionsmechanismus weisen zwei Kationen, **Cu^{2+}** und **Hg^{2+}** , auf: Sie binden an die Imidazolstickstoffatome von His64, dem *proton shuttle*, und blockieren somit dessen Funktion des intramolekularen H^+ -Transfers. Dadurch wird die katalytische Aktivität der CA herabgesetzt. Es handelt sich also um **Inhibitoren des Protonen-Shuttles** [Eriksson et al., 1988; Magid, 1967]. Für Cu^{2+} wurde gegen CA II ein IC_{50} von $0.5 \mu\text{M}$ bei pH 7.3 bestimmt.

Auch **Hydroxamate** besitzen eine inhibitorische Wirkung auf CAs. Für die beiden einfachen Verbindungen Acetohydroxamsäure und Trifluoracetohydroxamsäure wurden die IC_{50} -Werte zu $\text{IC}_{50}(\text{CH}_3\text{CONHOH}) = 47 \mu\text{M}$ und $\text{IC}_{50}(\text{CF}_3\text{CONHOH}) = 3.8 \mu\text{M}$ gegen CA II bestimmt [Scolnick, 1997]. Die Kristallstrukturanalyse des Komplexes von Acetohydroxamsäure und CA II zeigt einen den Sulfonamiden analogen Binde-modus (Abb. 3.7); allerdings wird das Zinkion nicht – wie man es bei Hydroxamaten und Thermolysin oder Matrixmetalloproteinasen beobachtet – bidentat von beiden

Tabelle 3.2: Inhibitionskonstanten einiger anionischer CAI gegen CA I, II und IV.
(nach [Supuran et al., 2004])

Anion	$K_i(\text{CA I})[\text{mM}]$	$K_i(\text{CA II})[\text{mM}]$	$K_i(\text{CA IV})[\text{mM}]$
F^-	>300	>300	–
Cl^-	6	200	36
Br^-	4	63	52
I^-	0.3	26	11
CNO^-	0.0007	0.03	0.03
SCN^-	0.2	1.6	–
N_3^-	–	1.5	–
HCO_3^-	12	85	44
HPO_4^{2-}	3	36	27
HCOO^-	–	20	6
CH_3COO^-	7	79	22
SO_4^{2-}	–	>200	44
ClO_4^-	3.6	1.3	–
NO_3^-	7	35	–

Sauerstoffatomen koordiniert. Stattdessen bildet das Carbonylsauerstoffatom der Hydroxamatgruppe eine Wasserstoffbrücke mit der Hauptketten-NH-Gruppe von Thr199 aus. Die Hydroxylgruppe derselben Aminosäure fungiert mit dem Thr199O γ als Wasserstoffbrückenakzeptor gegenüber dem Wasserstoffatom der OH-Gruppe der Hydroxamsäure. Das Wasserstoffatom der Thr199-OH-Gruppe dient als Donor gegenüber der Carboxylatgruppe von Glu106. Das deprotonierte, negativ geladene Stickstoffatom der Hydroxamsäuregruppe koordiniert in tetraedrischer Geometrie das Zinkion. Im Falle des Trifluoro-Derivates bildet ein Fluoratom eine zusätzliche schwache C-F \rightarrow Zn $^{2+}$ Wechselwirkung (2.8 Å) aus; diese Interaktion in Verbindung mit dem durch den starken Elektronenzug erniedrigten pK $_a$ -Wert der NH-Gruppe erklärt die deutlich höhere Affinität der Trifluoracetohydroxamsäure im Vergleich zur unsubstituierten Verbindung. Durch Einführung geeigneter Substituenten konnten Verbindungen mit nanomolaren Inhibitionskonstanten erhalten werden (z. B. die Serie von sulfonylierten Aminosäure-Hydroxamaten [Scozzafava & Supuran, 2000a]).

Eine weitere Klasse von CAI ohne Sulfonamidstruktur stellen **Sulfamate, Sulfamide und Hydroxysulfonamide** dar. Die kleinst möglichen Vertreter für die Sulfamate

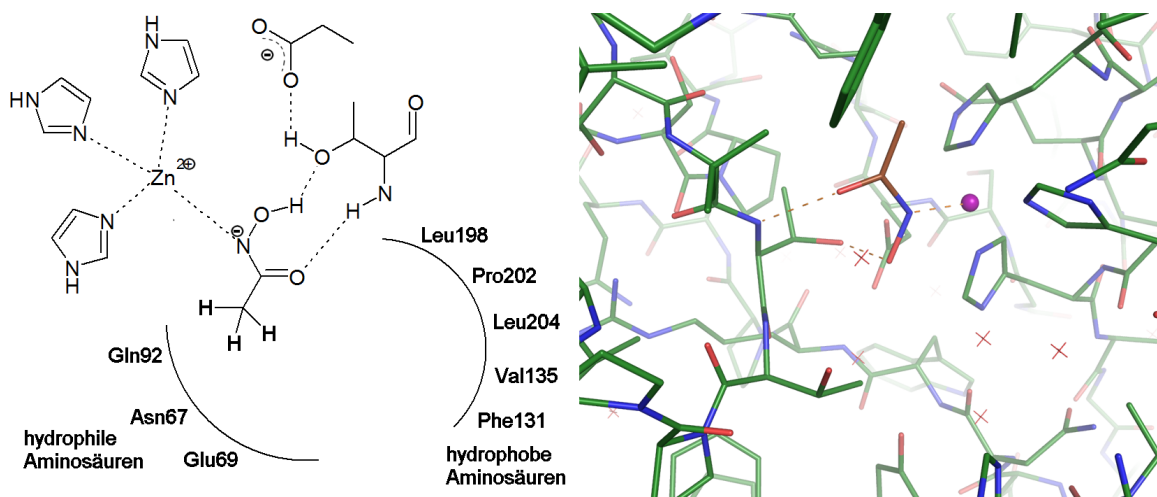


Abbildung 3.7: Bindungsmodus von Acetohydroxamsäure an CA II. Im linken Teil sind die Protein-Ligand-Wechselwirkungen schematisch dargestellt. Der rechte Teil zeigt die dreidimensionale Darstellung des Komplexes von Acetohydroxamsäure (Kohlenstoffatome braun) mit CA II (Kohlenstoffatome grün) (PDB-code: 1AM6). Das Zinkion ist als violette Kugel dargestellt, Wassermoleküle als rote Kreuze und Wasserstoffbrücken als gestrichelte Linien.

und Sulfamide, Sulfamid und Sulfaminsäure, zeigen schwache Inhibition von CA II ($K_i(H_2NSO_2NH_2) = 1.13 \text{ mM}$; $K_i(H_2NSO_2OH) = 390 \mu\text{M}$ [Briganti et al., 1996]); dennoch konnte für beide eine Kristallstruktur im Komplex mit CA II aufgeklärt werden [Abbate et al., 2002]. Die hierbei beobachtete Koordination ist analog der von unmodifizierten Sulfonamiden und wird bei den organischen Sulfamaten/-amiden in derselben Form gefunden: Ein deprotoniertes Stickstoffatom koordiniert das Zinkion, während ein Wasserstoffatom an demselben Stickstoff als Donor eine Wasserstoffbrücke zum Thr199O γ ausbildet. Ein Sulfonylsauerstoffatom dient als Akzeptor einer Wasserstoffbrücke, die von der Hauptketten-NH-Gruppe des Thr199 ausgeht. Das andere Sulfonylsauerstoffatom bildet einen schwachen Kontakt zum Zn^{2+} aus. Durch entsprechende Substitution konnten auch von dieser Verbindungsklasse nanomolare Inhibitoren erhalten werden. Das prominenteste Sulfamat ist das Antiepileptikum **Topiram** (Tab. 3.3, 7 auf Seite 50 und 51).

Die größte Klasse der CAI stellen die **Sulfonamide**. Sie weisen ein hochkonserviertes, typisches Wechselwirkungsmuster auf: Das aktivierte Wassermolekül und ein weiteres, das sog. „*deep water*“, werden verdrängt, die normalerweise primäre Sulfonamidgruppe liegt deprotoniert vor und koordiniert mit dem Stickstoffatom an das Zinkion. Das noch

vorhandene H-Atom an der NH-Gruppe bildet als Donor eine Wasserstoffbrücke mit dem Thr199O γ aus. Ferner fungiert dieselbe Hydroxylgruppe des Thr199 als Donor einer Wasserstoffbrücke zur Carboxylatgruppe des Glu106. Ein Sulfonylsauerstoffatom der Sulfonamidgruppe ist Akzeptor einer Wasserstoffbrücke, die von der Hauptketten-NH-Gruppe des Thr199 ausgeht, das andere bildet einen schwachen Kontakt zum Zinkion aus (meist etwa um 3 Å). Der an die Sulfonamidgruppe angefügte Molekülteil beeinflusst durch weitere Wechselwirkungen die Affinität des Sulfonamids. Im Falle der CA II besteht die Bindetasche etwa aus einer hydrophilen Hälfte (Aminosäuren Gln92, Asn67, Glu69) und einer hydrophoben (Phe131, Val135, Leu204, Pro202, Leu198). Beispielsweise bilden Inhibitoren mit einem Thienothiopyran-Gerüst mit ihrer zweiten, endozyklischen Sulfongruppe als Akzeptor eine zusätzliche Wasserstoffbrücke mit dem Gln92N ϵ aus, was sie zu hochpotenten Inhibitoren macht (Abb. 3.8) [Smith et al., 1994]. Ein Pentafluorobenzoylderivat des Methazolamids bildet eine Wasserstoffbrücke mit Gln92N ϵ mit Hilfe der Carbonylgruppe als Akzeptor aus. Der perfluorierte Phenylring nimmt eine parallele Position zum Phenylring des Phe131 ein und bildet eine π - π -stacking-Wechselwirkung aus [Abbate et al., 2003].

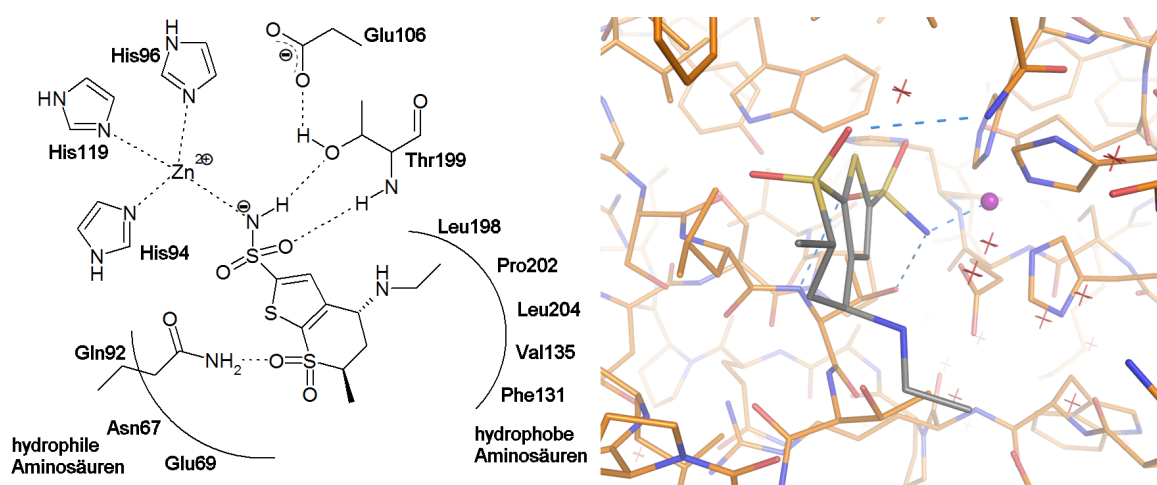


Abbildung 3.8: Bindungsmodus von Dorzolamid an CA II. Im linken Teil sind die Protein-Ligand-Wechselwirkungen schematisch dargestellt. Der rechte Teil zeigt die dreidimensionale Darstellung des Komplexes von Dorzolamid (Kohlenstoffatome grau) mit CA II (Kohlenstoffatome orange) (PDB-code: 1CIL). Das Zinkion ist als violette Kugel dargestellt, Wassermoleküle als rote Kreuze und Wasserstoffbrücken als gestrichelte Linien.

Aufgrund des allgemeinen Bindemodus der Sulfonamidgruppe führt eine Substituti-

on der primären, terminalen Sulfonamidgruppe am Stickstoff meist zum Verlust der inhibitorischen Wirkung [Krebs, 1948]. Daher ist es umso erstaunlicher, dass **Saccharin** (Tab. 3.3, **15**) ein Inhibitor der CAs ist. Zwar hemmt es die normalerweise für Sulfonamidinhibitoren hochsuszeptible CA II relativ schwach ($K_i = 5.8 \mu\text{M}$), zeigt aber als weitere Besonderheit einen hohen Grad an Selektivität (z.B. $K_i(\text{CA VII}) = 10 \text{ nM}$) [Köhler et al., 2007]. Im Unterschied zu den üblichen Sulfonamid-CAIs ist die Sulfonamidgruppe endozyklisch und zu einer Carbonylgruppe benachbart, so dass ein Sulfinimidsystem entsteht und den pK_a -Wert der NH-Gruppe auf 1.5 herabsetzt.

3.3.2 Therapeutisch genutzte Substanzen mit CA-inhibitorischer Wirkung

Es sind bereits mehrere Substanzen mit CA-Hemmwirkung im klinischen Gebrauch. Die Inhibition der CAs kann dabei entweder das pharmakologische Wirkprinzip für die Hauptwirkung sein, sie kann mehr oder weniger neben wichtigeren anderen Effekten die Hauptwirkung unterstützen, sie kann Ursache von unerwünschten Nebenwirkungen sein, oder sie ist lediglich *in vitro* feststellbar, ohne dass eine pharmakologische Relevanz (bisher) bekannt ist.

Abbildung 3.9 (S. 50) zeigt die Strukturformeln von therapeutisch eingesetzten Substanzen mit CAI-Wirkung, in Tab. 3.3 (S. 51) sind die entsprechenden INN und Inhibitionsdaten angegeben.

Acetazolamid (1) ist einer der ältesten verwendeten CAI. Hier ist die Hemmung der CAs Ursache für die Hauptwirkung; die Verwendung als Diuretikum ist heutzutage aufgrund der Nebenwirkungen systemisch angewendeter nicht-selektiver CAI (gastrointestinale Störungen, Müdigkeit, Appetitlosigkeit, Depressionen, Nierensteinbildung, metabolische Azidose, Störungen des Elektrolythaushaltes, metallischer Geschmack) obsolet. Acetazolamid wird eingesetzt bei Glaukom (nur kurzfristig oder bei Therapieresistenz gegenüber lokalen Antiglaukomatosa), speziellen Epilepsieformen, Höhenkrankheit oder benigner intracranieller Hypertension (erhöhter Schädelinnendruck ohne Zellwucherungen) sowie Hirnödem. **Methazolamid (2)** ist etwas lipophiler als Acetazolamid, sonst aber weitgehend mit diesem vergleichbar.

Ein ähnliches Indikationsspektrum besitzt **Ethoxzolamid (3)**, ein Benzothiazolsulfonamid. Es wird zur Behandlung von Glaukom, Duodenalkarzinom, als Diuretikum

und bei speziellen Epilepsieformen verwendet. Ein weiteres oral appliziertes Antiglaukomatosum ist das **Dichlorphenamid** (4). Es wird jedoch nicht zur langfristigen Monotherapie angewendet. Chemisch ist es ein dichlorsubstituiertes Phenylsulfonamid mit zwei meta-ständigen Sulfonamidgruppen.

Einen wesentlichen Fortschritt für die Glaukomtherapie stellte die Einführung zweier lokal angewandter Substanzen, **Dorzolamid** (5) und **Brinzolamid** (6), dar. Beide Moleküle sind ausreichend hydrophil, um als Augentropfen formuliert werden zu können, und ausreichend lipophil, um die Hornhaut des Auges zu durchdringen. Durch die lokale Applikationsform sind sie praktisch frei von den typischen CAI-Nebenwirkungen.

Während bei den Substanzen 1–6 die Inhibition der CAs für die erwünschte Hauptwirkung verantwortlich ist, ist sie bei den neurologisch eingesetzten Arzneistoffen 7–10 entweder ein untergeordneter Teil des Wirkmechanismus oder Ursache unerwünschter Nebenwirkungen. Das Sulfamat **Topiramat** (7) wird bei bestimmten Epilepsieformen und zur Migräneprophylaxe eingesetzt. Als Hauptwirkmechanismus wird eine Blockade des AMPA-Rezeptors, die Potenzierung der GABA-Wirkung am $\text{GABA}_A\text{-Cl}^-$ -Kanal und die Inaktivierung spannungsabhängiger Natriumkanäle angesehen. Prinzipiell ist eine epileptische Krampfneigung bei Azidose verringert, was die Wirkung von Acetazolamid bei Epilepsie erklärt. Also könnte die CA-Inhibition synergistisch mit den Effekten an den Ionenkanälen wirken, ihr wird jedoch eine untergeordnete Rolle zugeschrieben. Die vermehrte Nierensteinbildung als Nebenwirkung ist aber vermutlich auf die CA-Hemmung zurückzuführen. **Zonisamid** (8) ist ebenfalls zur Zusatztherapie bei Epilepsie zugelassen. Die Hauptwirkung wird hier durch allosterische Verstärkung der GABA-Wirkung, Inaktivierung spannungsabhängiger Na^+ -Kanäle und Reduktion des Ca^{2+} -Einstroms durch T-Typ-Calciumkanäle erklärt. Eine äußerst interessante Nebenwirkung von Topiramat und Zonisamid ist der beobachtete Gewichtsverlust. Dieser ist möglicherweise auf Inhibition von CA II, VA und VB (siehe Kap. 3.2.2 und 3.2.5) zurückzuführen [Simone & Supuran, 2007]. Für Zonisamid wird derzeit untersucht, ob dessen Anwendung in Verbindung mit einer hypokalorischen Diät als Antiadipositas-Therapie in Frage kommt. Über den Wirkmechanismus des Antiepileptikums **Sulthiam** (9) ist wenig bekannt. Allerdings scheint bei dieser Verbindung die CA-Inhibition (v.a. CA II und VII) maßgeblich an der antikonvulsiven Wirkung beteiligt zu sein [Temperini et al., 2007].

Die folgenden Arzneistoffe inhibieren CAs in signifikantem Ausmaß, jedoch ist die pharmakologische Relevanz dieser Beobachtung ungeklärt oder gering. **Sulpirid** (10) ist ein

Antidepressivum und atypisches Neuroleptikum. Diese Indikationen sind in der antagonistischen Wirkung auf Dopamin-D₂/D₃-Rezeptoren begründet. **Furosemid** und **Hydrochlorothiazid** (**11**, **12**) sind potente Schleifendiuretika zur Therapie von Hypertonie, Herzinsuffizienz und Ödemen. Die diuretische Wirkung rührt aber hauptsächlich von einer Hemmung der Na⁺- und K⁺-Rückresorption im aufsteigenden Schenkel der Henleschen Schleife her. Die Hemmung der CA trägt allerdings zusätzlich zur Ausscheidung von Bicarbonat bei.

Celecoxib und **Valdecoxib** (**13**, **14**) gehören zur Klasse der COX-2-selektiven NSARs. Beide wurden zur Behandlung von rheumatoider Arthritis, Osteoarthritis und Dysmenorrhoe angewendet. Valdecoxib wurde 2005 von *Pfizer* zurückgezogen aufgrund erhöhter kardiovaskulärer Risiken. Celecoxib wird ferner bei Morbus Bechterew und Familiärer Adenomatöser Polyposis (FAP) verwendet. Aufgrund der primären Sulfonamidgruppen hemmen beide Substanzen CAs. Es konnte gezeigt werden, dass beide Substanzen nach systemischer Applikation an Hasen den Augeninnendruck senken. Ferner wird vermutet, dass die bei *in vitro* und *in vivo* Studien beobachteten antineoplastischen Eigenschaften nicht nur auf die COX-2-Inhibition, sondern zumindest teilweise auf CA-Hemmung (v.a. CA II, IX, XII) zurückzuführen sind [Dogné et al., 2007].

Eine kürzlich beschriebene Kuriosität ist die Hemmwirkung von **Saccharin** (**15**) auf CAs [Köhler et al., 2007]. Es besitzt keine pharmakologische Wirkung, wird aber in großem Maßstab als Geschmackskorrigenz in verschiedenen Arzneiformen und als Süßstoff in der Lebensmittelindustrie benutzt. Es fällt insofern aus der Reihe der üblichen CAI mit terminaler, primärer Sulfonamidgruppe als hier die Sulfonamidgruppe endozyklisch und in ein Sulfimid eingebaut ist. Dies erniedrigt den pK_a-Wert der NH-Gruppe erheblich, allerdings ist kein zweites Wasserstoffatom mehr am Stickstoff, welches normalerweise eine Wasserstoffbrücke zum Thr199O γ ausbildet. Ferner zeigt Saccharin ein außergewöhnliches Selektivitätsprofil. Dieses hängt wahrscheinlich mit der Carbonylgruppe zusammen, die bei einigen Isoenzymen zu sterischen Konflikten mit voluminöseren Aminosäuren (z. B. CA I: His200) führt, bei anderen wiederum über ein Wassermolekül einen zusätzlichen Wasserkontakt – unter der Voraussetzung, dass geeignete Aminosäuren in der Bindetasche vorhanden sind (Ser, Thr) – mit dem Protein vermittelt. Da Saccharin bisher selbst in hohen Dosen beim Menschen keine Nebenwirkungen gezeigt hat, scheint die CA-Hemmung in den für die Substanz zugänglichen Kompartimenten keine Relevanz zu besitzen. Allerdings könnte der unangenehme, bittere metallische Nachgeschmack eine Folge der Inhibition der im Speichel enthaltenen

CA VI sein. Solche Geschmacksirritationen werden häufig bei klinisch angewendeten CAI beobachtet und sind medizinisch bedeutungslos, sie sind aber im Hinblick auf den Einsatz von Saccharin als Geschmacksverbesserer durchaus von Interesse.

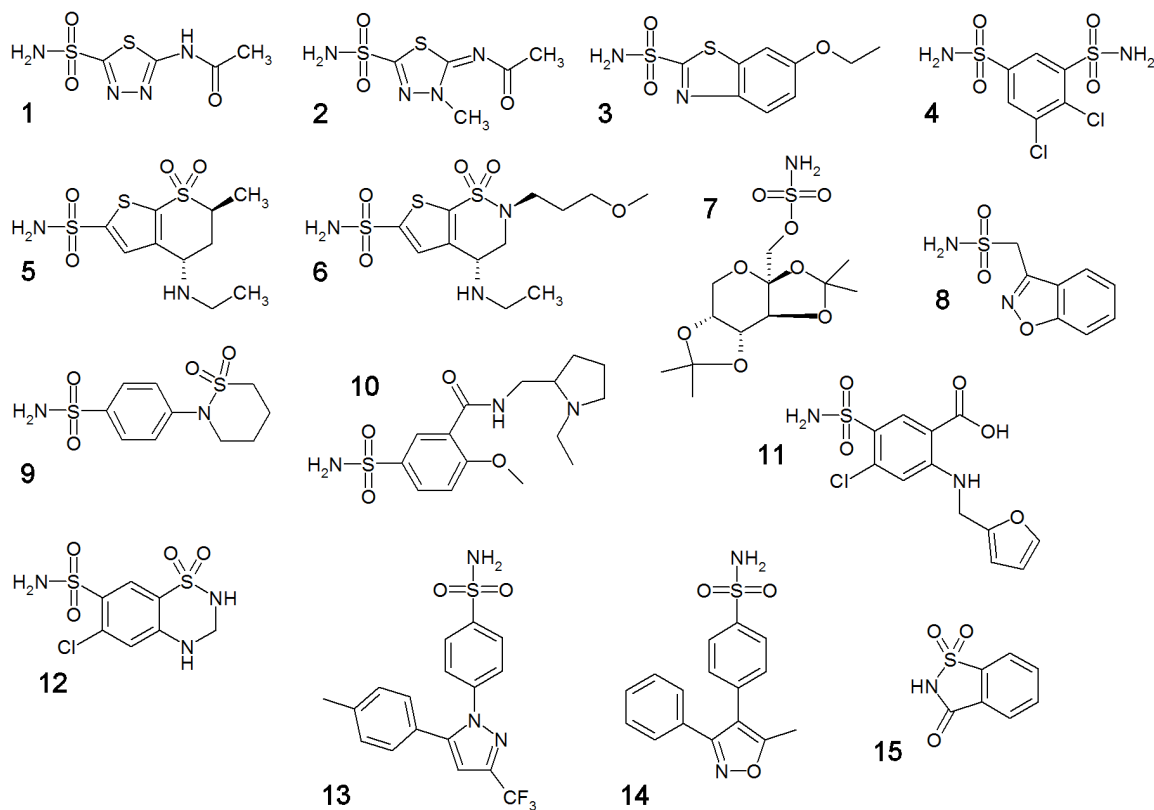


Abbildung 3.9: Strukturformeln von therapeutisch genutzten Substanzen mit CA-inhibitorischer Wirkung. Der INN der Substanzen und die entsprechenden K_i -Werte sind in Tab. 3.3 angegeben.

Tabelle 3.3: Inhibitionsdaten therapeutisch verwendeter Substanzen mit CA-inhibitorischer Wirkung. Die Strukturformeln sind in Abb. 3.9 aufgeführt (Daten aus [Dogné et al., 2007; Nishimori et al., 2007, 2005b; Abbate et al., 2004; Temperini et al., 2007; Köhler et al., 2007]).

K _i [nM] gegen CA-Isoenzym ^{a)}													
Nr.	Name	I	II	III	IV	VA	VB	VI	VII	IX ^{d)}	XII ^{d)}	XIII ^{b)}	XIV
1	Acetazolamid	250	12	$2 \cdot 10^5$	74	63	54	11	2.5	25	5.7	17	41
2	Methazolamid	50	14	$7 \cdot 10^5$	6200	65	62	10	2.1	27	3.4	19	43
3	Ethoxzolamid	25	14	$1.1 \cdot 10^6$	93	25	19	43	0.8	34	22	50	25
4	Dichlorphenamid	1200	38	$6.8 \cdot 10^5$	15000	630	21	79	26	50	50	23	345
5	Dorzolamid	50000	9	$7.7 \cdot 10^5$	8500	42	33	10	3.5	52	3.5	18	27
6	Brinzolamid	45000	3	$1.1 \cdot 10^5$	3950	50	30	0.9	2.8	37	3.0	10	24
7	Topiramamat	250	10	$7.8 \cdot 10^5$	4900	63	30	45	0.9	58	3.8	47	1460
8	Zonisamid	56	35	$2.2 \cdot 10^6$	8590	20	6033	89	117	5.1	11000	430	5250
9	Sulthiam	374	7	$6.3 \cdot 10^5$	95	81	91	134	6	43	56	1450	1540
10	Sulpirid	12000	40	n.b.	620 ^{c)}	174	18	0.8	n.b.	46	n.b.	n.b.	n.b.
11	Furosemid	62	65	n.b.	564	499	322	245	513	420	261	550	52
12	Hydrochlorothiazid	328	290	n.b.	427	4225	603	3655	5010	367	355	3885	4105
13	Celecoxib	50000	21	n.b.	290 ^{c)}	n.b.	n.b.	n.b.	n.b.	16	n.b.	n.b.	n.b.
14	Valdecoxib	54000	43	n.b.	340 ^{c)}	n.b.	n.b.	n.b.	n.b.	27	n.b.	n.b.	n.b.
15	Saccharin	18540	5950	n.b.	7920	10060	7210	935	10	103	633	12100	773

n.b. = nicht bestimmt

^{a)}soweit nicht anders angegeben, gegen das vollständige humane, rekombinante Enzym gemessen

^{b)}murines Enzym

^{c)}bovines Enzym

^{d)}katalytische Domäne

4 3D-QSAR-Studien zur Identifizierung affinitäts- und selektivitätsbestimmender Faktoren

Dieses Kapitel ist bereits als Artikel [Hillebrecht et al., 2006] publiziert worden. Deshalb erscheint es als Bestandteil dieser Doktorarbeit in englischer Sprache. Der Inhalt entspricht der Originalpublikation; es wurde jedoch auf den *Abstract* und die *Introduction* verzichtet, da die methodische Einführung bereits in Kap. 2 erfolgt ist. Geringfügige Adaptionen sind ansonsten nur vorgenommen worden, soweit sie die Formatierung betreffen.

4.1 Einleitung

Die folgenden Abschnitte beschreiben die Erstellung, Validierung und Interpretation verschiedener 3D-QSAR-Modelle (CoMFA, CoMSIA, AFMoC) an den drei Isoenzymen CA I, II und IV. Die Isokonturkarten der rein ligandbasierten Modelle (CoMFA, CoMSIA) wurden denen des rein proteinbasierten GRID/CPCA-Ansatzes gegenübergestellt und mit den Ergebnissen des „gemischten“ AFMoC-Verfahrens verglichen.

Weiterhin wurden die statistischen Parameter der Modelle quantitativ verglichen sowie der Einfluß des Alignments in den verschiedenen Proteinbindetaschen der Isoformen auf die Ergebnisse untersucht.

4.2 Target Proteins

This study deals with the human carbonic anhydrase (hCA) isoforms carbonic anhydrase I, CA II, and CA IV, which consist in their mature forms of 260, 259, and 266 amino acids, respectively. CA I and II are cytosolic enzymes, while CA IV is bound to the plasma membrane. All three isozymes possess a zinc ion, coordinated by three histidine residues His94, His96, and His119 (amino acid numbering in the following

according to hCA I). The fourth ligand is a water molecule that must be deprotonated for the generation of the catalytically active form. It further hydrogen bonds to the hydroxyl group of Thr199. This threonine residue itself forms a hydrogen bond to the neighboring carboxylate moiety of Glu106. The substrate, carbon dioxide, is located in a hydrophobic pocket close to Zn(II), but not coordinated to it [Lindskog, 1997; Christianson & Cox, 1999; Lindskog & Coleman, 1973]. The three conserved histidine residues and zinc together with Thr199 build the basis for the binding of sulfonamides and similar substructures (sulfamates, sulfamides) [Abbate et al., 2002]: The deprotonated inhibitors coordinate to zinc via a negatively charged nitrogen [Lindskog, 1997], the remaining hydrogen acting as hydrogen-bond donor to the hydroxyl group of Thr199. One of the sulfon oxygens accepts a hydrogen bond from the NH-group of Thr199, whereas the other forms an additional polar interaction to the zinc ion. This general binding geometry remains virtually unchanged for the closely related sulfamates (scaffold **I**, Fig. 4.1), sulfamides (**H**), or hydroxysulfonamides (**F**). The crystal structures of all three human isozymes have been determined at sufficiently high resolution, and a C α superimposition of the proteins reveals a rather unchanged geometry of the binding pocket, with the exception of His64, which acts as the “proton shuttle” and adopts two alternative conformations [Tu et al., 1989, 1998]

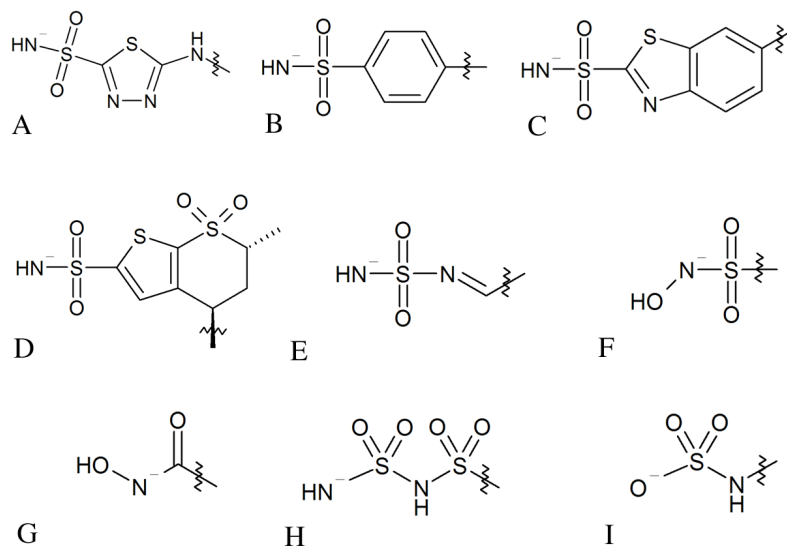


Figure 4.1: Nine different scaffolds contained in the ligand sets, used for 3D-QSAR studies.

4.3 Methods

All molecular modeling and QSAR studies were performed using SYBYL 7.0 [SYBYL 7.0] (CoMFA and CoMSIA) and DrugScore 1.2 (AFMoC). Chemometric analyses were carried out using GOLPE 4.5 [Clementi, 1999]. Several analyses and validation processes were automated using SPL (SYBYL Programming Language) and the scripting language Python. All figures were generated using PyMOL 0.98 [DeLano, 2005].

4.3.1 Dataset and Alignment

A suitable dataset for QSAR has to be structurally sufficiently diverse, however, still representative to provide the desired scope of the derived models [Golbraikh et al., 2003; Clark, 1997, 1998]. If no common pattern is shared by the considered molecules, it might be impossible to extract a correlation between the descriptors and the biological activities. This would also make the alignment problem difficult. In a previous study [Weber, 2004; Weber et al., 2006], we used 87 inhibitors to construct affinity and selectivity models for CA I, II, and IV. However, some of the derived models showed only q^2 values at the border of significance to properly model pK_i differences. Therefore, we tried to profit in the present study from an enlarged data basis and compiled a better distributed and balanced data set. For QSAR studies, the affinity data should spread over a range of at least three logarithmic units [Thibaut et al., 1993] and display approximately an even distribution. In the present study, these requirements are particularly difficult to fulfill, since we have to deal with six dependent variables simultaneously (three pK_i values for the three isozymes and three pairwise differences of the pK_i values (*vide infra*)). From literature a data set of initially 1748 compounds was selected. These were reduced to a structurally diverse set of 173 ligands [Supuran et al., 1999; Supuran, 1999; Menabuoni et al., 1999; Popescu et al., 1999; Scozzafava & Supuran, 1999; Scozzafava et al., 2000c; Supuran & Scozzafava, 2000a; Supuran, 1998; Ilies et al., 2000; Casini et al., 2000, 2002; Scozzafava et al., 2001; Mincione et al., 2001; Masereel et al., 2002; Supuran et al., 2000b; Renzi et al., 2000; Mincione et al., 1999; Supuran et al., 2000a; Barboiu et al., 2000; Scozzafava et al., 1999; Borrás et al., 1999; Supuran & Scozzafava, 2000c; Scozzafava & Supuran, 2000b; Scozzafava et al., 2000b; Supuran et al., 1998; Scozzafava & Supuran, 2000a; Scozzafava et al., 2000a] using the SELECTOR module of SYBYL (Tanimoto similarities of 2D and atom pair fingerprints). It comprises nine different scaffolds (see Fig. 4.1). To check the informa-

tion content of the data with respect to selectivity studies, the correlation coefficient of the affinity data was calculated: A Pearson correlation coefficient of 0.5 for CA I/II and 0.57 for CAI/IV was obtained. For the combination of CA II/IV, however, a rather high degree of correlation ($r = 0.86$) indicates that any subsequent conclusion concerning this isozyme pair has to be treated with particular caution.

In 3D-QSAR studies, a meaningful alignment of the ligands is of utmost importance. Many different solutions of this problem have been reported, considering either the ligands or proteins as the basis of the alignment rule [Klebe, 1993; Lemmen & Lengauer, 2000]. Since crystal structures of the proteins under consideration are available, we decided to manually place the molecules into the binding pockets. The X-ray structures 1azm (CA I), 1cil (CA II), and 1znc (CA IV) were retrieved from the Protein Data Bank (PDB) [PDB-RCSB; Berman et al., 2000] and superimposed by fitting the C_α atoms of conserved amino acids, using the BIOPOLYMER module of SYBYL. Subsequently the ligands and the water molecules were removed with the exception of the four water molecules W301, W325, W394, and W395 in 1cil, since they will most likely remain in the active site upon ligand binding and their consideration in docking has proved beneficial in earlier studies [Grüneberg et al., 2002]. Coordinates of the training set ligands were converted from 2D to 3D using the program CORINA [Gasteiger et al., 1990]. Protonation states were assigned as follows: Carboxylic acids were assumed to be deprotonated whereas non-aromatic amines and amidino groups were considered as protonated. In addition to these standard assumptions, also the primary sulfonamide or sulfonate group was assigned to the deprotonated state, as previously evidenced by NMR [Lindskog, 1997]. After calculation of Gasteiger-Marsili charges [Gasteiger & Marsili, 1980] the molecules were minimized using the TRIPOS force field [Clark et al., 1989] (maximum number of iterations: 1000, minimum gradient: 0.005, default values for the remaining parameters). The thus obtained ligand geometries were matched onto corresponding atoms in the crystal structure references. Once the anchor fragment was superimposed onto the crystal coordinates, the remaining torsions of the ligand were adjusted to achieve maximal overlap to an appropriate crystal structure template. Finally, to relax the preplaced ligand into an appropriate force-field minimum the MAB force field [Gerber & Müller, 1995] as implemented in MOLOC was used with default settings. During minimization the protein and the zinc anchor of the ligand were kept fixed. Fig. 4.2 shows the alignment of the training set ligands in CA II.

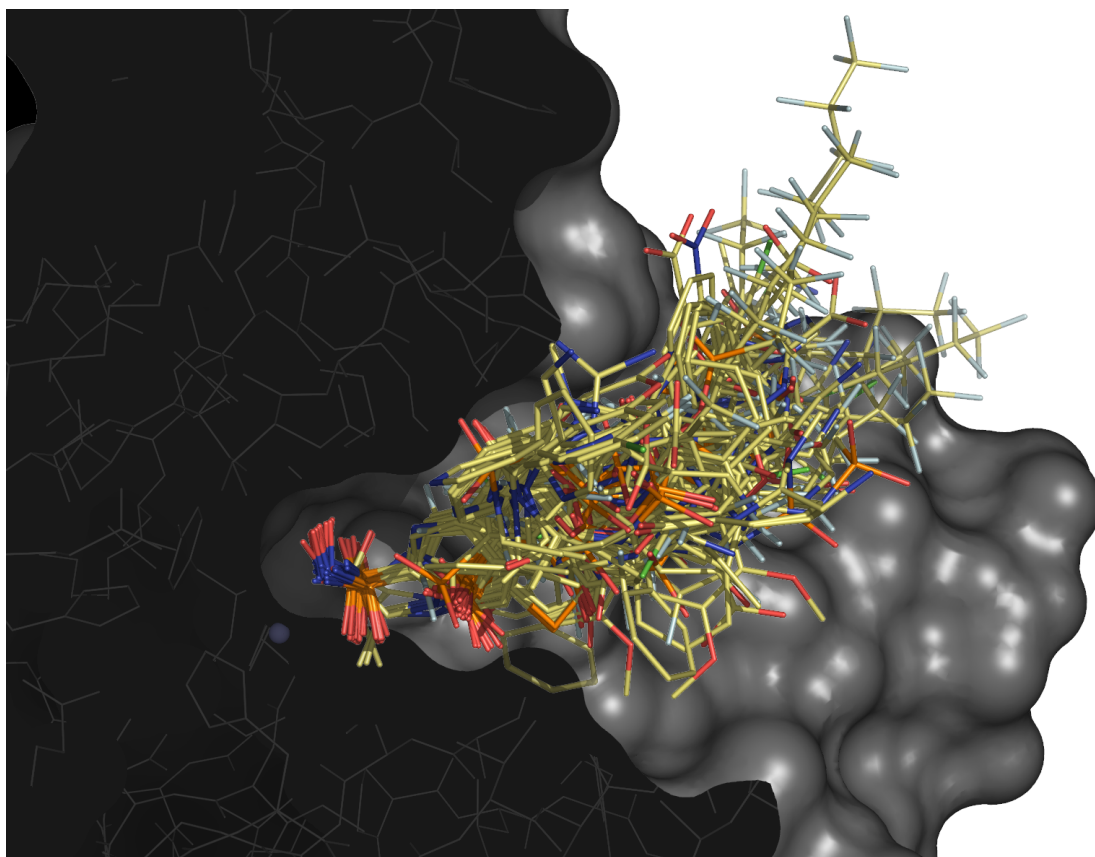


Figure 4.2: Alignment of the inhibitors used for 3D-QSAR studies, shown together with the binding site of CA II. The solvent accessible surface is indicated as a solid surface.

4.3.2 CoMFA and CoMSIA Analyses

For CoMFA analysis the interaction energies between a probe atom and the ligand atoms were calculated using a grid box of $26 \times 34 \times 25$ points with 1 Å spacing, embedding all ligands with a margin of at least 4 Å in each direction. The same box dimensions were also used for CoMSIA, GRID and AFMoC studies. A positively charged sp^3 -carbon atom was used as the probe atom for calculating steric and electrostatic CoMFA fields applying SYBYL standard parameters (TRIPOS standard field, dielectric constant $1/r$, cutoff 30 kcal/mol). CoMSIA fields were computed for steric, electrostatic, hydrophobic, and hydrogen bonding properties, using a probe of charge +1, radius of 1, hydrophobicity and hydrogen bonding properties of +1, and an attenuation factor α of 0.3 for the Gaussian distance-dependent function. All fields were scaled with the CoMFA-STD scaling procedure, assigning equal weights to each field. For each of the three alignments affinity values (or affinity differences, respectively)

were correlated with the field descriptors using SAMPLS [Bush & Nachbar, 1993] in a leave-one-out cross-validation analysis. Ligands associated with severe uncertainties during alignment or with major singularities in structural terms (especially the class of hydroxamates) were removed from the set, resulting in a collection of 144 ligands. After the optimal number of components had been determined, a PLS analysis was performed without cross-validation, applying no column filtering.

In order to further assess the statistical significance of the generated models, we exemplarily performed the following analyses based on the alignment in CA II: A more rigorous cross-validation was carried out by leaving groups of up to 10% of the molecules out, in order to test the stability of the models. Furthermore, all possible field combinations were considered as descriptor variables in various SAMPLS runs in order to check the importance of each field type. To assess whether the models are significant and not the result of chance correlations, the affinity (selectivity) data were scrambled randomly [Cramer et al., 1993] and subsequently the models were recalculated. This procedure was repeated 100 times for each dependent variable and mean q^2 values were calculated over all runs; this was done for each of the 10 extracted components.

Finally, the data set was divided into a training set of two third and a test set of one third of all ligands. The separation was done using complete linkage clustering of the data set into 48 structurally diverse clusters based on a Tanimoto similarity measure calculated for 2D and atom pair fingerprints. From each cluster one compound was randomly selected and assigned to the test set, while the remaining molecules were assigned to the training set. After generation of PLS models from the training set, the dependent variables of the test set ligands were predicted and a predictive r^2 was calculated.

4.3.3 AFMoC Analysis

DrugScore pair potentials were calculated for the binding pockets of the three isozymes, using the PDB structures 1azm (CA I), 1cil (CA II), and 1znc (CA IV), respectively. Location, spacing, and orientation of the grid box as well as the individual ligand alignments were the same as used for CoMFA/CoMSIA analyses. Subsequently, interaction fields were calculated by multiplying the DrugScore potentials with weights derived from Gaussian functions (width $\sigma = 0.85$), which were centered at the ligand atom positions. This procedure was performed for eight atom types available in DrugScore:

C.3, C.2, C.ar, N.3, N.ar, N.am, O.2, and O.3 (atom type notation similar to the SYBYL atom type conventions). The decision whether to include an individual atom type into the analyses or not was based on the frequency of occurrence of the respective type in the data set, the supposed importance for the protein-ligand interactions, the statistical significance of this atom type in the data set used to derive DrugScore potentials, and some preliminary AFMoC models, based on selected combinations of atom types (data not shown). A value of 10 was taken for the height of the Gaussian repulsive term at short atom-atom distances. The thus obtained interaction fields were correlated with the experimental affinity (differences) using the SAMPLS routine implemented in AFMoC. Statistical performance is assessed by similar parameters as in the other 3D-QSAR techniques with the difference that two corresponding sets of parameters are assigned: one refers to the original experimental pK_i value ($pK_{i,tot}$), the second to a slightly modified pK_i value, $pK_{i,PLS}$. The $pK_{i,PLS}$ value is obtained by subtracting the mean expected affinity contribution from atom types not considered in the PLS analysis from the original experimental pK_i value. This contribution is estimated according to the original DrugScore potentials by applying the empirically derived scaling factor $c_S^{PAIR} = -1.25 \cdot 10^{-5}$. In total, 167 molecules were used for the generation of the AFMoC models.

As for CoMFA/CoMSIA, some additional statistical tests were accomplished exemplarily for the alignment in CA II. A y-scrambling test was performed in the same way as for CoMFA/CoMSIA, and a leave-five-out analysis was carried out to test the stability of the models. For all three alignments, the data set was divided into training and test set analogously to the procedure described above for CoMFA/CoMSIA. The obtained predictive r^2 values served as measure for the predictive power of the AFMoC models.

4.3.4 3D-QSAR Based Selectivity Maps

In order to visualize the selectivity determining regions in space, two different options are available, as described above: One is to contour the $stdev \cdot coeff$ maps of the models obtained by employing pK_i differences of two isozymes as independent variables. The second approach is to subtract the $stdev \cdot coeff$ values at corresponding grid points from each other after individual models for two isozymes have been generated. In case of AFMoC the second strategy allows to generate individual models each using the protein corresponding to the respective pK_i value. After PLS analysis and subtraction of

corresponding $\text{stdev} \times \text{coeff}$ the resulting maps contain selectivity information not only via the affinity values but also by evaluating the two deviating protein environments. Since in the case of the second strategy the models have been built separately, the scale of the $\text{stdev} \times \text{coeff}$ might differ, especially if the similarity of the compared targets is not considerably large. Therefore, it could be advantageous to normalize the values separately before the subtraction. We tried both approaches, and in the case of normalization, we scaled each $\text{stdev} \times \text{coeff}$ field to values between -1 and 1. All herein described techniques to obtain selectivity maps were applied to CoMFA, CoMSIA, and AFMoC, respectively, and compared to each other.

4.3.5 Protein Based Selectivity Maps

The above-mentioned methods utilize either purely ligand based information (CoMFA, CoMSIA) or a mixture of both, protein and ligand information (AFMoC). A third possibility is to extract knowledge solely from protein structures. The protein structures were superimposed onto 1cil with SYBYL using the C_α atoms, and all water molecules and heteroatoms except zinc ions were removed. Only the structures of human wildtype proteins were included in the analyses, namely: 1azm, 1bzm, 1czm, and 1hcb for CA I, 1bcd, 1bn1, 1cil, and 1g1d for CA II, and 1znc for CA IV, respectively. The conformation of the most flexible residue, His64, was modeled as the “out”-conformation in order to avoid the generation of artefacts at this position. If available, we included several X-ray structures per isozyme, to assess the robustness and sensitivity of the methods. As the first step, a grid box of the same dimensions and location as for the 3D-QSAR studies was placed into the binding pockets of all three isozymes. Subsequently, the program GRID [Goodford, 1985] was used to calculate the interaction energies between four selected probes and the protein atoms treating non-hydrogen receptor atoms as rigid (directive MOVE = 0). These probes represent four important functionalities of protein ligand interactions: The DRY probe (hydrophobic, entropic contributions), the C3 probe (hydrophobic, steric contributions), the N1 probe (neutral planar NH, e.g. amide, donor function), and the O probe (sp^2 -carbonyl oxygen, acceptor function). To extract information from the obtained descriptors relevant for selectivity between two proteins, we tested two approaches: The first simply subtracts the energy values of two corresponding grid points from each other and the resulting differences are used for the contour maps. The second is to perform a consensus principal component analysis (CPCA) [Kastenholz et al., 2000] as implemented into the program GOLPE [Clementi,

1999]. The calculated GRID fields were imported into GOLPE and pretreated in the following way: First a maximum cutoff of 0 kcal/mol was applied in order to focus on favourable interactions only, following the recommendations of several authors [Kastenholtz et al., 2000; Naumann & Matter, 2002]. Then, all data points with an absolute value smaller 0.01 kcal/mol or a standard deviation below 0.02 kcal/mol were set to zero. Furthermore, 2- and 3-level variables (i.e. variables taking on only two or three distinct values over the entire matrix) were discarded and the “cut-out tool” was used to remove all data points from the following analyses which had a distance beyond 4.5 Å from an artificial “ligand”, which was constructed in a way to sufficiently cover the entire binding sites of the enzymes. Block unscaled weights (BUW) were applied in order to give all four probes the same initial importance in the CPCA models. Three principal components were extracted, and so-called “active CPCA differential plots” were visualized and interpreted for each of the four probes. This type of plots allows the user to draw a vector between two objects in the scores plot for a distinct probe. GOLPE displays in an associated plot the selectivity fields calculated from the differences of the scores of the objects at the ends of the drawn vector.

4.4 Results and Discussion

4.4.1 Statistical Results

The molecular alignments of the inhibitors generated for the three different isozymes were rather similar. Figure 4.2 shows the alignment of the inhibitors in the cavity of CA II. Supposedly, in this enzyme the most reliable superposition could be achieved compared to the other two isozymes, since 60 CA II crystal structures with bound inhibitors were available as reference, while for the other two cases only three (CA I) or one (CA IV) templates could be extracted from the PDB.

Table 4.1 on page 76 shows the statistical results for all CoMFA, CoMSIA, and AFMoC analyses with individual pK_i values as independent variables using all three separate alignments. The q^2 values of all models are greater than 0.3 indicating their statistical significance [Cramer et al., 1993]. Moreover, almost all q^2 are above 0.5, except for the CoMFA and AFMoC models for CA I, if the alignment generated in CA I or CA IV is used. In general, the prediction of the pK_i value of CA I seemed to be affected by larger residual errors than for CA II and IV. Most of the remaining models show

excellent statistical significance, indicated by q^2 values above 0.6. Interestingly, the ligand alignment in CA II exhibits the highest q^2 values for most of the cases, irrespective which set of pK_i values is predicted. Several factors might explain this finding: The ligand alignment in CA II is supported by a large number of crystal structures, whereas for the others only a few reference structures are available. Furthermore, the finally accepted superposition in each binding pocket revealed only minor differences. In general, it has been suggested that 3D-QSAR models perform better if the ligands are superimposed in a highly consistent way [Oprea & Waller, 1997; Klebe & Abraham, 1993]. As a more rigorous test of the stability of our models, a cross-validation with leave-10%-out was performed, where the data set is divided into ten groups of equal size. Each group is left out once during cross-validation, and the q^2 is calculated. If the obtained q^2 values significantly decrease in comparison to those of a leave-one-out-analysis (LOO-analysis), the models are probably instable (as obviously the large LOO- q^2 is only revealed due to a high degree of redundancy in the data set) [Doweyko, 2004]. The medium decrease in q^2 for CoMFA was 0.02, for CoMSIA 0.01, confirming the robustness of the models. For AFMoC, a leave-five-out analysis was carried out yielding average differences between LOO- q^2 and leave-five-out- q^2 of 0.02 for both, q_{tot}^2 and q_{PLS}^2 .

The statistics for the models obtained with pK_i differences are shown in Tab. 4.2 (p. 77). For these models in all cases q^2 values above 0.5 could be obtained, demonstrating the significance also of the selectivity models. The dependency of the performance on the alignment is small, as for the individual affinity values, and the alignment in CA II yields in almost all cases the highest q^2 , too. Nevertheless one has to bear in mind that the dependent variable in this case is a difference of two individual variables and is therefore affected by an error in the magnitude of the sum of the single variables' errors, due to the laws of error propagation.

We also examined another approach to predict selectivity, where the differences between pK_i values are calculated from the pK_i predictions of the individual models. The statistical parameters are shown in Tab. 4.3 (p. 78). Compared to the approach using the pK_i differences as independent variable, in most cases this method performs virtually identically well. The difference in the respective q^2 of the two approaches is mostly below 0.05 except for the AFMoC selectivity models for CA II and CA IV. For all 3D-QSAR techniques used herein this selectivity property was the most difficult to predict. This is not unexpected due to the high initial correlation of the target-specific

pK_i values.

As outlined by many authors, the predictive r^2 value is the most informative parameter to assess the external predictive power of QSAR models [Golbraikh et al., 2003; Oprea & Waller, 1997; Doweiko, 2004; Golbraikh & Tropsha, 2002]. It reflects best the “real life” situation if compounds are predicted which were not included in the process of model derivation. However, in our studies the trends revealed by the q^2 are also present in the predictive r^2 values, as shown in Tab. 4.1, 4.2, and 4.3 (on pages 76 to 78). The CoMSIA approach showed best performance in almost all cases, while CoMFA performed slightly worse. For the major part of pK_i (difference) predictions, predictive r^2 values clearly above 0.5 could be achieved, finally proving the predictive power of the models. In general, poorest predictivity was encountered for the models of CA I affinities and the selectivity between CA II and IV, respectively. AFMoC was also able to produce robust estimates of most of the dependent variables, though with decreased performance compared to the other two 3D-QSAR approaches. The optimum mixing factor for each model was determined to be 1.0 in all cases. The reduced performance of AFMoC compared to the purely ligand-based approaches may be explained by the huge amount of ligand information provided by the large size of the data sets. Obviously, these data sufficiently cover the relevant aspects of the underlying structure-activity (-selectivity) relationships, such that the incorporation of protein information does not improve the predictive power of the models. This is also in accordance with the finding that a mixing factor of 1.0 is optimal for all models, meaning that a maximum of ligand information yields the best results.

In order to assess that no chance correlations were responsible for the results, a y-randomization test was carried out exemplarily for the alignment produced in CA II, because it yielded the best models. The mean q^2 values, averaged over 100 runs, were negative without exception, confirming the relevance of our models.

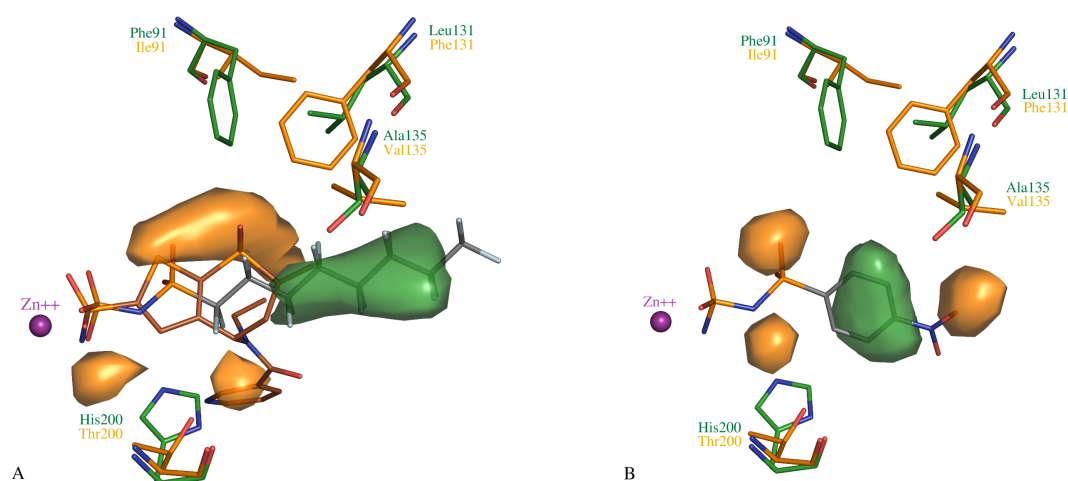
4.4.2 Visualization of the Property Maps

One prominent feature of 3D-QSAR methods is the possibility to visualize the regions in space that are responsible for an increase or decrease of a particular type of dependent variable, thus providing additional information about the structure-activity and -selectivity relationship. Contours are drawn around those grid points, where the coefficient of the QSAR equation multiplied with the standard deviation of the corre-

sponding descriptor column ($\text{stdev} \times \text{coeff}$) is above or below a user defined threshold. Some of the maps reflect properties mirrored by the receptor, but it has to be pointed out that this is not necessarily the case, since – for CoMFA and CoMSIA – only ligand information is evaluated to derive the models or – as in the present study – also for AFMoC plays the dominant role due to the size and scope of the data set. All CoMSIA maps shown are generated by PLS analysis of the pK_i differences, using the alignment in CA II. The AFMoC selectivity maps were obtained performing PLS analyses for each pK_i value separately using the corresponding protein and calculating the differences of the QSAR coefficients afterwards. Also in this case only the ligand alignment in CA II was used.

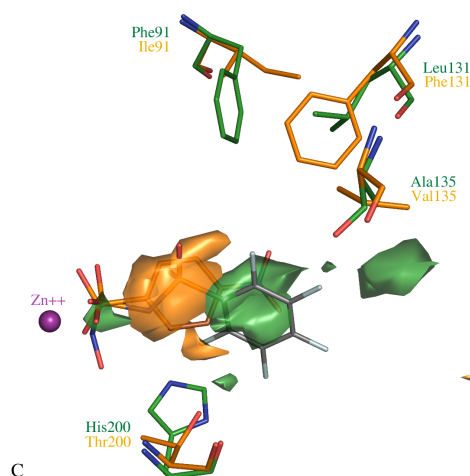
Figure 4.3(a) shows the steric CoMSIA contour map of the selectivity model, built on the differences of the pK_i values of CA I and CA II. The catalytic zinc ion (magenta sphere) as well as selected amino acid residues of CA I (green) and CA II (orange) are depicted. The green isopleth indicates a region where steric bulk increases the selectivity towards CA I, while sterically demanding substituents occupying the orange regions increase selectivity towards CA II. Additionally, the inhibitors **16** (Fig. 4.4) ($\Delta \text{pK}_i(I - II) = -2.60$, brown) and **17** ($\Delta \text{pK}_i(I - II) = 0.86$, grey) are displayed as examples for ligands with a preference to inhibit CA I or CA II, respectively. It has to be noted that the “preference” for a distinct isozyme is only considered relatively to the range and distribution of the respective pK_i difference. E.g. a compound (like **17**) with a $\Delta \text{pK}_i(I - II)$ of 0.86 cannot be termed “highly selective” for CA I from a medicinal chemistry point of view, but in comparison to other entries in the dataset, it shows a higher tendency to bind to CA I. The green contour is mostly occupied by molecular portions present in scaffold **H** (see Fig. 4.1). This compound class shows highest preference for the isozyme CA I, and most of the CA I selective regions are associated with that class. In CA I, the residues Leu131 and Ala135 open up a larger space for sterically demanding moieties compared to Phe131 and Val135 in CA II. The large orange contour in Figure 4.3(a) indicates a CA II selective region, where especially the oxygen atoms of the endocyclic sulfon group of the thienothiopyranes (scaffold **D**, e.g. compound **16**) is frequently found. The active site of CA II exhibits here an isoleucine residue (Ile91), whereas in CA I a phenylalanine reduces the available space. The two smaller orange contours can be related to the exchange of His200 (CA I) by Thr200 (CA II). Some of the CA II selective thienothiopyranes (e.g. **16**) possess larger substituents at their amino group, which point into the right contour.

The hydrophobic CoMSIA contours are shown in Figure 4.3(b) together with compound



(a) Steric CoMSIA map. Green contours (level: 0.002) indicate regions enhancing selectivity towards CA I, orange contours (level: -0.0005) show areas enhancing selectivity towards CA II. For reference, inhibitors **16** ($\Delta pK_i(I - II) = -2.60$, brown) and **17** ($\Delta pK_i(I - II) = 0.86$, grey) are depicted.

(b) Hydrophobic CoMSIA map. Green contours (level: 0.004) indicate regions enhancing selectivity towards CA I, orange contours (level: -0.0033) show areas enhancing selectivity towards CA II. For reference inhibitor **18** (grey) is depicted.



(c) Combined C.2, C.3, and C.ar AFMoC map. Green contours (levels: C.2: -0.008, C.3: -0.015, C.ar: -0.017) indicate regions favouring CA I selectivity, orange contours (levels: C.2: 0.015, C.3: 0.009, C.ar: 0.025) depict areas increasing selectivity towards CA II. For reference the inhibitors **20** (grey) and **19** (brown) are shown.

Figure 4.3: Comparison of CoMSIA and AFMoC $\text{stdev} \times \text{coeff}$ contour maps for CA I/II selectivity. The catalytic zinc ion (violet sphere) is shown as well as selected amino acids of CA I (green) and CA II (orange); hydrogens are omitted, for clarity.

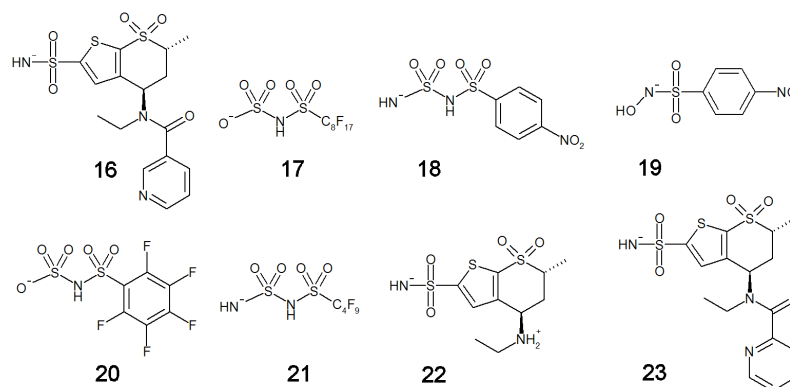
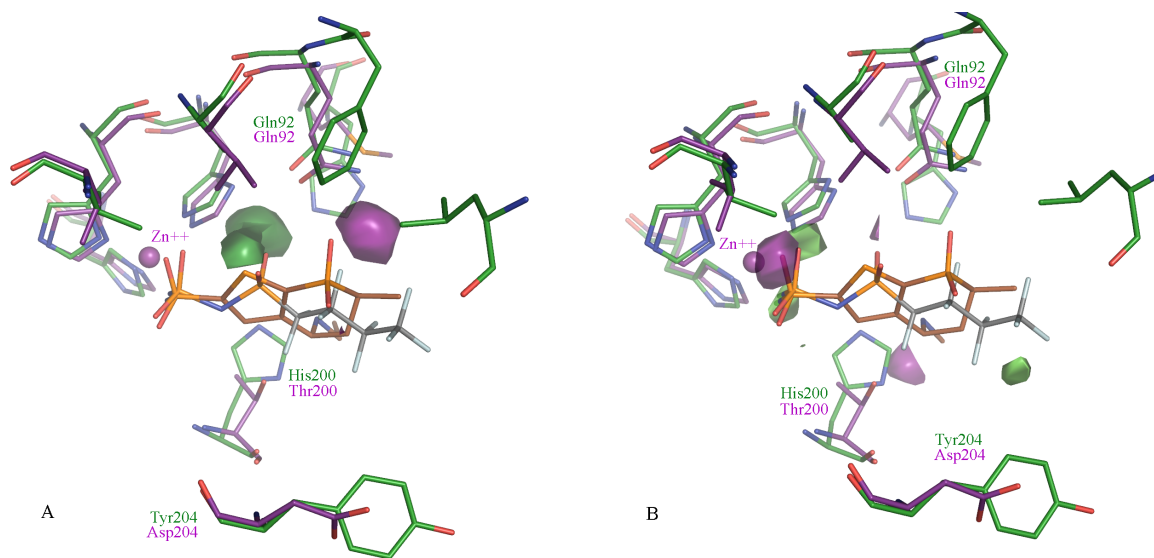


Figure 4.4: Chemical formulae of the molecules mentioned in the Results and Discussion section.

18 ($\Delta pK_i(I-II) = 0.60$, grey) using the same color coding. In this case no obvious correlation of the amino acid composition of the active site residues and the presence of these contours can be established, presumably because entropic and desolvation effects play a predominant role for binding (e.g. green contour next to the phenyl ring of **18**). Such properties can be important for ligand binding and require hydrophobic regions on the ligand surface, however they do not necessarily reflect intimate correspondence on the side of the protein binding pocket. Also the orange contour, enclosing the second sulfon group not being part of the zinc anchor, reflects the preference of this inhibitor class for CA I: The presence of hydrophobic moieties in that area enhances CA II selectivity, while hydrophilic moieties provoke CA I selectivity. One of the remaining orange contours marks a position where several ligands with scaffold **H** possess polar groups, such as the oxygen atoms of the nitro group of **18**. Obviously such groups are detrimental for CA II binding but enhance selectivity towards CA I.

The steric and hydrophobic property maps were selected in order to compare the CoMSIA contours to the corresponding AFMoC maps. As the AFMoC approach is based on atom types rather than on generic properties, the fields of several atom types corresponding best to the considered CoMSIA properties have to be combined to enable such a comparison. In the present case we merged the C.3, C.2, and C.ar fields. They are displayed in Figure 4.3(c) together with the hydroxysulfonamide **19** ($\Delta pK_i(I-II) = -3.80$, brown) and compound **20** ($\Delta pK_i(I-II) = 1.38$, grey), applying the same color coding as before. The larger green contour around the phenyl ring of **20** coincides well with the respective hydrophobic CoMSIA contour (Fig. 4.3(b)). AFMoC marks a region occupied by the phenyl ring of **19** as CA II selective which is

in accordance with the steric and hydrophobic CoMSIA maps at that location. The small orange fragment close to His200 also finds its counterpart in the steric CoMSIA map (Fig. 4.3(a)).



(a) CoMSIA stdev*coeff acceptor map for CA I/IV selectivity. Green contours (level: 0.005) indicate regions enhancing selectivity towards CA I, purple contours (level: -0.0025) show areas increasing selectivity towards CA IV.

(b) AFMoC stdev*coeff O.2 map for CA I/IV selectivity. Green contours (level: -0.0025) indicate regions enhancing selectivity towards CA I, purple contours (level: 0.003) encapsulate areas increasing selectivity towards CA IV.

Figure 4.5: Comparison of CoMSIA and AFMoC stdev*coeff contour maps for CA I/IV selectivity. The catalytic zinc ion (violet sphere) is shown as well as selected amino acids of CA I (green) and CA IV (purple); hydrogens are omitted for clarity. For reference, inhibitors **21** ($\Delta pK_i(I - IV) = 1.15$, grey) and **22** ($\Delta pK_i(I - IV) = -3.05$, brown) are depicted.

As another example for mutually corresponding CoMSIA and AFMoC maps, the selectivity models between CA I and CA IV will be discussed, considering properties (or atom types, respectively), which are related to directed interactions. Figure 4.5 presents the acceptor CoMSIA (Fig. 4.5(a)) map and the O.2 AFMoC (Fig. 4.5(b)) map together with the catalytic zinc ion and inhibitors **21** ($\Delta pK_i(I - IV) = 1.15$, grey) and **22** ($\Delta pK_i(I - IV) = -3.05$, brown). Green contours highlight regions where the presence of an acceptor (or O.2 atom) enhances CA I selectivity, magenta contours indicate areas where the occurrence of an acceptor (or O.2 atom) increases selectivity towards CA IV. As in the previous case, it has to be noted that the correspondence of the selected maps of both approaches is not perfect. Also other atom types beside O.2

(e.g. O.3, N.ar) can act as acceptors, but O.2 is the most frequently occurring atom type with clear monofunctional acceptor properties in this data set. The green contour (Fig. 4.5(a)) close to the sulfon oxygens of **21** is present in the CoMSIA map only and can be attributed to a CA I preference found for scaffold **H**. However, these oxygens do not form hydrogen bonds to CA I. Since CoMSIA only evaluates ligand information, it simply reflects the observation that ligands with acceptors in this region possess CA I preference in a phenomenological way. On the other hand the AFMoC approach also exploits protein information and therefore it reveals no predominant contour, because the protein environment does not favour the presence of an O.2 atom in this region (Fig. 4.5(b)). The magenta contour in the CoMSIA map depicts a CA IV selectivity enhancing region. This area is close to the oxygen of the exocyclic sulfonamide group of the CA IV selective thienothiopyranes (e.g. **22**). In both proteins, CA I and CA IV, however, the conformation and intramolecular hydrogen bond network of the neighboring Gln92 prevents the formation of a hydrogen bond to the ligand oxygen. In consequence, AFMoC does not highlight this region by a corresponding contour.

A similar behaviour of the AFMoC method can be observed also in the other maps. It is interesting to note that some of the AFMoC maps show very sharply bordered, fragmented contours, which surround single atoms of ligands. Obviously in this region the respective dependent variable adopts rather extreme values. This possibly indicates a too steep functional form of the AFMoC potentials and minor alignment differences with respect to these atoms are highlighted as significant if the corresponding ligands differ in their selectivity profiles. One possibility to overcome this problem may be the choice of a value greater than 0.85 (σ) for the width of the Gaussian function that maps the ligand atom types onto the distance dependent pair potentials. This would lead to a broader “smearing” of the ligand atom derived weights and thus produce smoother gradients in the descriptor values.

As already described above, we investigated two alternative approaches to generate selectivity models for the three 3D-QSAR methods used. One is to correlate the pK_i differences (ΔpK_i) as independent variable with the ligand or interaction fields, respectively. The other is to build separate models for two pK_i values and to calculate the differences of corresponding PLS coefficients subsequently. In latter case, however, it has to be taken into account that due to the separate derivation of the PLS models the scale of the two subtracted coefficient maps might differ. Hence, we calculated the difference maps on the unscaled as well as on normalized maps, fitted to a value range between -1 and 1, and compared the results. Visual inspection shows that – at

least in our case – the normalization does not significantly alter the contours for any of the applied QSAR techniques. The same holds true for a comparison of the maps obtained by separate PLS evaluation and a single correlation with the pK_i differences for CoMFA and CoMSIA. Due to methodological reasons in the case of AFMoC, the “separate” variant allows to include information of two different protein structures and shows significant differences compared to the use of the ΔpK_i . This will be discussed in detail below.

4.4.3 Differential GRID Maps and CPCA

At first the protein-based contouring using GRID probes will be discussed. Three principal components (PC) were extracted during the derivation of the CPCA models. Table 4.4 (p. 79) shows the amount of variance captured in each PC on the consensus level (corresponding to an ordinary PCA with all four probes simultaneously) and on the block level (referring to a single probe). For all probes the original data structure can sufficiently be summarized by three PCs, in the case of the DRY probe, however, only approximately 55% of the variance (versus 72% for C3, N1, or O) can be explained by the first three PCs, indicating a higher degree of complexity in the underlying data structure.

Figure 4.6 shows the CPCA scores (block scores) plot of the O probe for the first two PCs. The overall PCA scores (consensus scores) plot as well as the block scores for the other three probes display almost identical distribution. The scores plot shows a well defined clustering of the representatives of the three isozymes. The first PC separates the CA I class from CA II and CA IV, while the second PC discriminates CA IV from CA I and CA II. This evaluation was compared with a simple subtraction of the corresponding GRID maps.

As an example, Figure 4.7 shows the O probe selectivity maps for CA I/CA II selectivity (active differential CPCA plot, loadings). In Figure 4.7(a) regions associated with CA I selectivity are depicted together with compound **23** ($\Delta pK_i(I - II) = -2.60$, brown) for orientation. The blue meshes were derived by the CPCA approach, whereas the green solid contours were generated by building the differences of the GRID values. The region close to His67 (CA I) is marked by both approaches as important for CA I selectivity. Three amino acid mutations may explain the occurrence of this contour: The nitrogen atoms of His67 and His200 in CA I point directly to this area and can

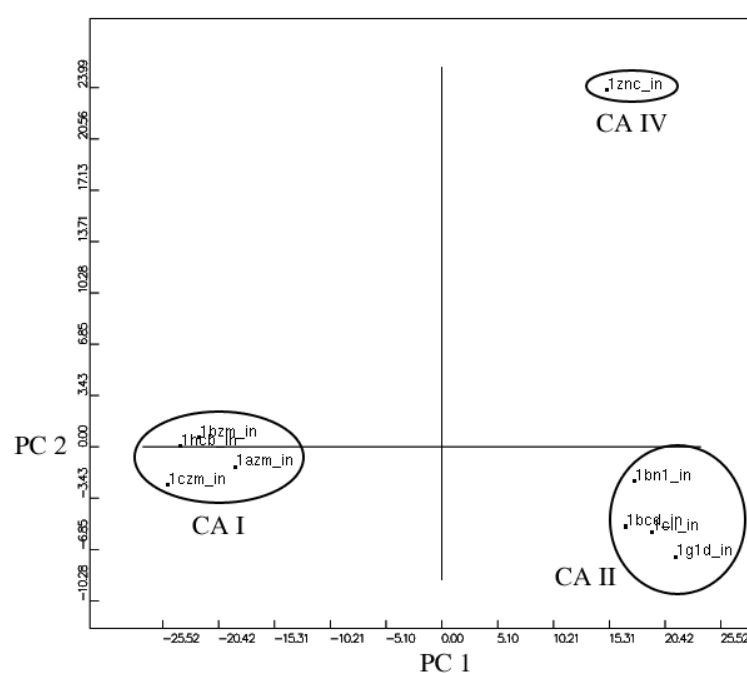
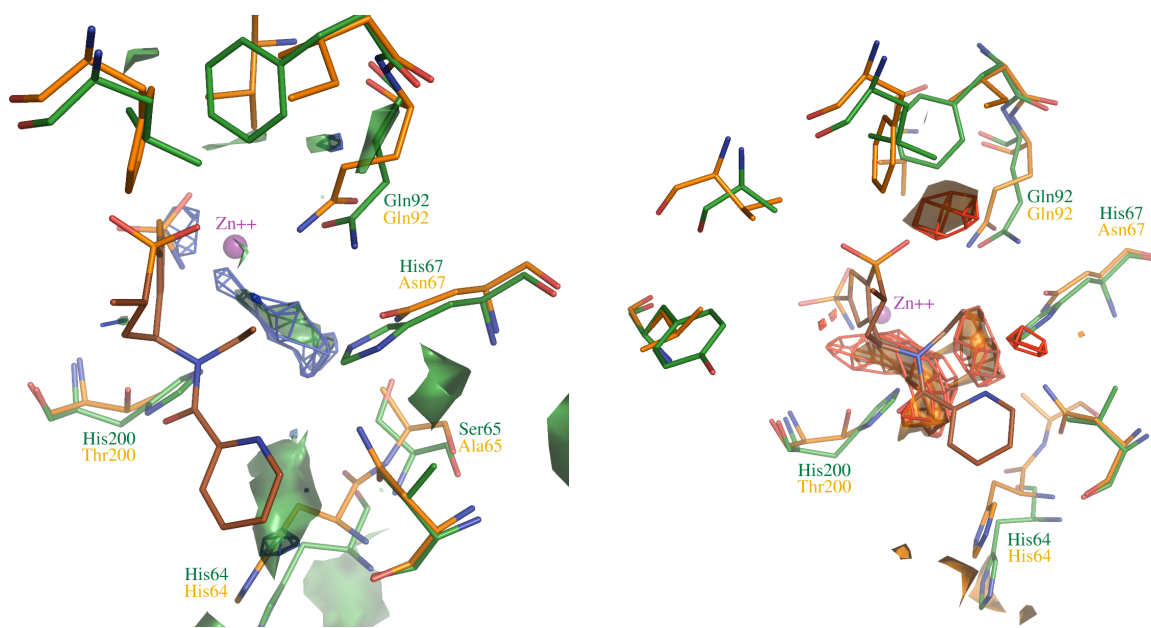


Figure 4.6: **CPCA scores plot for the carbonyl oxygen probe.** The clustering of the structures representing each isozyme is evident. PC 1 separates the CA I class from CA II and IV, whereas PC 2 discriminates CA IV from CA I and II.

act as donors if a suitable tautomeric form is assumed. The respective amino acid side chains of CA II (Asn67 and Thr200) also possess donor functions, but neither the angles nor the distances to the contour would allow them to build hydrogen bonds to a carbonyl oxygen placed at this region. Furthermore, Ser65 in CA I is close to the contour while CA II lacks a similar donor functionality in a residue side chain (Ala65). The green contour next to His64 (both isozymes) is obviously a result of alignment differences. This residue (“proton-shuttle”) is known for its conformational flexibility. Thus, it differs from structure to structure and the corresponding area should be excluded from selectivity considerations. The CPCA approach, however, seems to be less sensitive to such effects, as it indicates only a very small field contribution in this region. It has to be kept in mind, though, that the “region cut-out tool” of GOLPE was solely applied to the CPCA maps; consequently, contours in the periphery will appear in differential GRID maps only, because the corresponding variables have been discarded from the input matrix of the CPCA.



(a) CA I selectivity contour maps. Green solid contours (level: 7.0) were generated by the difference method, whereas blue meshes (level: -2.3) were derived by the CPCA approach.

(b) CA II selectivity contour maps. Orange solid contours (level: -8.0) were generated by the difference method, whereas red meshes (level: 2.3) were derived by the CPCA approach.

Figure 4.7: Protein derived CA I/II selectivity maps obtained by CPCA and subtraction of GRID descriptors. The catalytic zinc ion (violet sphere) and selected amino acids of CA I (green) and CA II (orange) are shown. For reference, inhibitor **23** ($\Delta pK_i(I - II) = -2.60$, brown) is depicted.

Figure 4.7(b) shows the regions attributed to CA II selectivity. The solid orange contours are produced by the GRID difference method, and red meshes are generated by CPCA. Both maps are almost identical. The large contour surrounding Thr200 (CA II) and His200 (CA I) is caused by the fact that in this area a hydrogen bond to the hydroxyl group of Thr200 is possible, whereas His200, whose nitrogen atom can also operate as donor, approaches too close to this region. A similar explanation holds for the neighboring contour associated with Asn67 (CA II) and His67 (CA I), respectively. The contour close to Gln92 is produced by the occurrence of different conformations of the same amino acid in both proteins. These conformations are induced by different amino acids in the vicinity of Gln92 and their respective involvement in a hydrogen bonding network to the interior of the protein: In CA I the nitrogen of Gln92 forms a hydrogen bond to Asn69 in the protein interior and, thus, does not point its donor functionality into the binding pocket. In contrast, in CA II the neighboring amino acid 69 is a glutamic acid that forms a bidentate salt bridge to Arg58, and it is hence unable to accept a hydrogen bond from Gln92. Consequently, Gln92 possesses an unsatisfied donor functionality that can be involved in binding. The corresponding carbonyl acceptor oxygen of the terminal amide group of Gln92 is engaged either in CA I and II in a hydrogen bond to the neighboring His94.

This comparison exemplarily shows that the use of differential GRID maps for the extraction of protein-based selectivity information is an alternative way to the well known CPCA approach. In our case, the maps of both methods show no significant differences. However, the advantage of CPCA is that it provides insight into the relationships between the studied proteins in terms of scores plots. Furthermore, it evaluates rather principle components than raw data, which should result in a reduction of noise likely contained in the descriptor matrix.

4.4.4 Synopsis of the Approaches

So far in this study, three methodologically different approaches have been presented to unravel selectivity determining features. In the following section, we will compare their results to feature the major differences.

Figure 4.8 shows the contour maps for CoMSIA (Fig. 4.8(a), acceptor field), AFMoC (Fig. 4.8(b), O.2 field, orange solid contours: separate PLS method using two proteins, red meshes: PLS based on ΔpK_i using CA II (1cil) as reference only). In Fig. 4.8(c)

differential GRID fields are presented (O field) to trace selectivity differences between CA I and II. For all contours the presence of the respective probe enhances selectivity for CA II in the highlighted region. For orientation, the CA II selective inhibitor **23** ($\Delta pK_i(I - II) = -2.60$, brown) is also depicted.

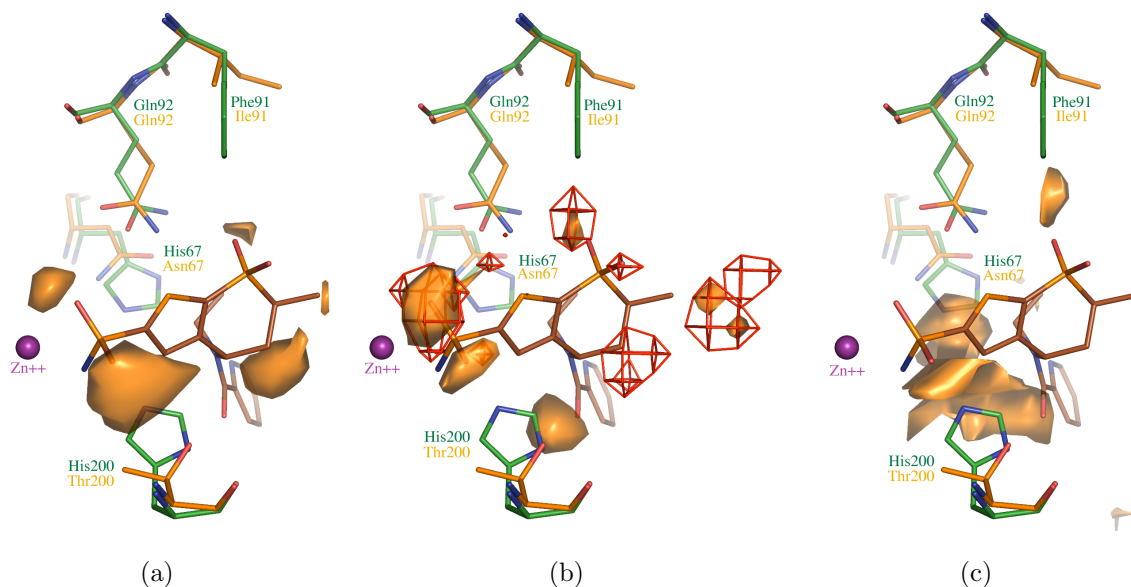


Figure 4.8: Synopsis of protein- and ligand-based CA I/II selectivity maps. The catalytic zinc ion (violet sphere) as well as selected amino acids of CA I (green) and CA II (orange) are depicted; hydrogens are omitted for clarity. For reference, inhibitor **23** ($\Delta pK_i(I - II) = -2.60$, brown) is shown. All depicted contours indicate areas enhancing selectivity for CA II. (a) CoMSIA stdev*coeff acceptor map (level: -0.0023). (b) AFMoC stdev*coeff O.2 maps. The solid orange contours (level: 0.002) were generated by building individual models for CA I and II and using both proteins. The red meshes (level: 0.002) were produced by PLS analysis with $\Delta pK_i(I - II)$ as dependent variable and only CA II as protein. (c) GRID differential O maps (level: -8.0), generated by subtraction of corresponding grid points without normalization.

The contours next to the zinc anchor, present either in the CoMSIA and AFMoC maps are difficult to interpret and probably result in consequence of slight alignment differences of the ligands. The large CoMSIA contour on the right of residue 200 (CA I: His, CA II: Thr) is evoked by the carbonyl or sulfonyl oxygen present in some potent para-amino substituted phenylsulfonamides showing CA II preference. This contour only reflects the empirically observed trend that ligands with an acceptor function at this position possess on average higher affinity for CA II, but no structural reason (e.g. a donor function in CA II which is not present in CA I) can be encountered in the receptor

structures. Accordingly, the purely protein-based differential maps produced by GRID do not show a CA II selectivity enhancing region in this area. In the case of AFMoC (Fig. 4.8(b)), the presence of a contour depends on the way the models were derived: If only one protein structure (CA II) is considered and PLS is performed with respect to ΔpK_i s as dependent variables (red meshes), the contour observed in CoMSIA is also indicated. If the map is constructed by evaluating the differences of the $stdev \cdot coeff$ from separately generated models including both protein structures (CA I and CA II), the region is not detected as significant. This can be explained by the fact that the former methodological variant considers the selectivity information through ligand information only (in terms of affinity differences data), whereas the latter approach additionally exploits the differences in the protein environments. Analogously, the red meshes to the very right in the AFMoC ΔpK_i map highlight an area where a few ligands with moderate CA II preference place their O.2 atoms. Since this trend is not very pronounced, the CoMSIA map (Fig. 4.8(a)) exhibits there only a very tiny contour. In the AFMoC contour map originating from two proteins (Fig. 4.8(b), solid contour), the contour is also strongly reduced taking into account that solely the ligand information substantiates the presence of such a contour while the protein structures show no obvious reason for it. Accordingly, also the differential GRID maps (Fig. 4.8(c)) lack a corresponding contour. The solid orange AFMoC contour in the direct vicinity (right) of His/Thr200 gives an example for the inverse case where the receptor environment produces a contour which is absent if ligand information is considered only. As indicated by the large orange differential GRID contour (Fig. 4.8(c)), the presence of an oxygen acceptor is favourable for CA II selectivity at this position due to the hydroxyl group of Thr200, which is replaced by a histidine residue in CA I. Several CA II selective inhibitors with a thienothiopyrane scaffold (e.g. **23**) possess an O.2 atom in this region. Obviously, this ligand information is not sufficient to produce a contour, since neither the CoMSIA nor the single protein AFMoC maps (red meshes) contour this region as important for CA II selectivity. However, the AFMoC approach using two protein structures (solid orange contour) seems to enhance the ligand-based information with additional knowledge about differences in the considered protein structures, thus producing that contour. A particularly interesting case are the contours around the endocyclic sulfon group of **23**. Considering only protein information the presence of one contour near Gln92 is expected, as already explained above. This contour is found in the maps of all approaches used herein. A contour around the second oxygen (not involved in hydrogen-bonding to Gln92) would be reasonable only from the point of view of the ligands, because all ligands with that thienothiopyrane structure exhibit

strong selectivity towards CA II. This difference can clearly be seen evaluating the maps of both AFMoC variants (Fig. 4.8(b)). The AFMoC model utilizing only one protein reveals red meshes at both oxygens, whereas the model including information of both proteins generates only one contour at the oxygen next to Gln92 which is in accordance with differences in the receptor structures (see above).

One area near His/Asn67 is contoured by the differential GRID approach only (Fig. 4.8(c)). Since this region is not occupied by any of the ligands in the considered dataset, none of the 3D-QSAR methods identifies this region as important for CA II selectivity. Therefore, it is always advisable to also consult purely protein-based approaches as supplement, if possible.

This comparison demonstrates that the AFMoC approach highlights regions as important for selectivity similar to those also marked by the ligand-based CoMSIA approach. However, the AFMoC contours are partly modified according to the differences in the receptor structures, either generating additional contours or diminishing the ligand-based contours.

4.5 Conclusions

The optimization of lead structures does not only require the enhancement in terms of their affinity; also the elaboration of a satisfactory selectivity profile is of utmost importance in order to avoid undesired side effects. In this study, we selected the carbonic anhydrases I, II, and IV as an example, as insufficient selectivity of inhibitors can produce serious problems due to pronounced similarity of these isozymes. A large high-quality dataset was used to derive statistically significant and predictive 3D-QSAR (CoMFA, CoMSIA) models to quantitatively estimate affinity and selectivity data. The interpretation of the contour maps is consistent with relationships found for spatial properties of the inhibitors and the trends of the data to predict. However, since these methods solely exploit ligand information, the contours resembled only in part the requirements and differences defined by the protein structures. To supplement the analyses also with protein information, the binding sites were examined with GRID on the one hand, and AFMoC models were generated on the other hand. The developed models can be used for further design in two ways: The contour maps provide hints for possible structural modifications, while the QSAR equations can be used to predict

the binding properties within inhibitor series, e.g. of focused combinatorial libraries, particularly if they are designed in a way that they do not extrapolate beyond the applicability domain defined by the employed models.

In summary, the following amino acid residues are most probably determinant to be considered for design of selective inhibitors: His200 is only present in CA I whereas a threonine constitutes the corresponding residue in CA II and IV. The introduction of sterically demanding substituents into potential ligands close to this residue will enhance selectivity towards CA II and IV. If the substitutions additionally exhibit an acceptor function that can form a hydrogen bond to the hydroxyl group of the threonine the selectivity increase can even be enhanced. To achieve an opposing selectivity profile the same histidine residue can be addressed by a sterically confined acceptor function. This group can further interact with His67 (CA II: Asn67; CA IV: Met67) and Ser65 (CA II: Ala65; CA IV: Ser65), present in CA I, to enhance selectivity towards CA I. Since CA IV, but not CA II (Ala65), possesses a serine at position 65 too, a similar substitution might be used to increase selectivity towards CA IV relatively to CA I or II. Furthermore, the static differences of the conformations of Gln92 due to deviating hydrogen-bond networks could be exploited for the design of CA II selective inhibitors by introduction of a suitable acceptor function. Eventually, also the steric differences next to the residues 91 and 131 could be exploited to favor selectivity either towards CA I or II. But due to the small distance between these residues and their inverse steric demands (CA I: Phe91/Leu131; CA II: Ile91/Phe131) an appropriate chemical modification of an inhibitor targeting only the properties of one of these residues will be difficult. Further options for the design of selective inhibitors can be deduced from the other contours discussed in the synopsis section, even if they are not directly linked to the differences of an amino acid exchange in the active sites. They reflect empirically correlations between structural differences of the ligands and their achieved affinities/selectivities. Thus they may provide valuable guidelines how to perform ligand modification, though not intimately explainable by one particular spot in the protein structures.

The comparison of the different contour maps underlines the complementary character of ligand- and protein-based approaches and demonstrates the capability of AFMoC to efficiently combine both sources of information.

Table 4.1: Statistical Results of CoMFA, CoMSIA, and AFMoC^{a)} Analyses of the Affinity Data Using Three Alignments.

	pK _i I			pK _i II			pK _i IV		
	CoMFA	CoMSIA	AFMoC ^{b)}	CoMFA	CoMSIA	AFMoC ^{b)}	CoMFA	CoMSIA	AFMoC ^{b)}
Alignment in CA I									
q^2	0.449	0.530	0.466 (0.423)	0.779	0.840	0.820 (0.806)	0.704	0.777	0.760 (0.752)
s_{PRESS}	0.789	0.728	0.808 (0.808)	0.615	0.522	0.542 (0.542)	0.684	0.591	0.605 (0.605)
r^2	0.848	0.862	0.651 (0.623)	0.922	0.918	0.913 (0.906)	0.907	0.895	0.863 (0.858)
S	0.414	0.395	0.653 (0.653)	0.366	0.374	0.378 (0.378)	0.384	0.405	0.458 (0.458)
F	154.45	172.23	75.38 (66.83)	325.63	387.31	278.60 (256.70)	268.03	297.67	202.50 (194.50)
# comp.	5	5	4	5	4	6	5	4	5
pred. r^2	0.345	0.508	0.180	0.746	0.816	0.781	0.735	0.776	0.652
Alignment in CA II									
q^2	0.530	0.580	0.492 (0.469)	0.853	0.860	0.791 (0.784)	0.825	0.819	0.755 (0.754)
s_{PRESS}	0.731	0.689	0.787 (0.787)	0.504	0.489	0.580 (0.580)	0.529	0.534	0.612 (0.612)
r^2	0.909	0.860	0.635 (0.618)	0.949	0.943	0.858 (0.853)	0.961	0.935	0.854 (0.854)
S	0.322	0.397	0.667 (0.667)	0.297	0.313	0.480 (0.480)	0.249	0.321	0.472 (0.472)
F	227.98	169.63	70.48 (65.50)	423.84	453.48	244.00 (234.50)	482.75	395.21	188.80 (187.80)
# comp	6	5	4	6	5	4	7	5	5
pred. r^2	0.423	0.499	0.219	0.767	0.825	0.658	0.772	0.779	0.663
Alignment in CA IV									
q^2	0.435	0.512	0.385 (0.367)	0.822	0.851	0.753 (0.743)	0.749	0.785	0.709 (0.705)
s_{PRESS}	0.798	0.742	0.864 (0.864)	0.552	0.504	0.631 (0.631)	0.630	0.581	0.665 (0.665)
r^2	0.845	0.851	0.526 (0.513)	0.935	0.931	0.825 (0.818)	0.920	0.920	0.790 (0.786)
S	0.418	0.410	0.758 (0.758)	0.333	0.343	0.531 (0.531)	0.356	0.355	0.565 (0.565)
F	150.39	157.38	60.39 (57.25)	398.01	466.89	191.10 (182.50)	317.50	397.61	151.90 (149.10)
# comp	5	5	3	5	4	4	5	4	4
pred. r^2	0.402	0.513	-0.091	0.838	0.833	0.590	0.792	0.784	0.450

^{a)}For calculation of DrugScore fields, the same protein as taken for the alignment was used.

^{b)}The first value refers to the PLS models using $pK_{i,tot}$, the second value in parentheses refers to $pK_{i,PLS}$ (for explanation, see text).

Table 4.2: Statistical Results of CoMFA, CoMSIA, and AFMoC^{a)} Analyses of the Selectivity Data Using ΔpK_i as Dependent Variable and Three Alignments.

	pK _i I – II			pK _i I – IV			pK _i II – IV		
	CoMFA	CoMSIA	AFMoC ^{b)}	CoMFA	CoMSIA	AFMoC ^{b)}	CoMFA	CoMSIA	AFMoC ^{b)}
Alignment in CA I									
q^2	0.734	0.744	0.601 (0.637)	0.725	0.711	0.628 (0.652)	0.546	0.606	0.532 (0.585)
s_{PRESS}	0.622	0.608	0.763 (0.763)	0.591	0.603	0.665 (0.665)	0.449	0.416	0.452 (0.452)
r^2	0.963	0.941	0.801 (0.819)	0.948	0.910	0.814 (0.826)	0.871	0.814	0.727 (0.758)
S	0.231	0.292	0.539 (0.539)	0.256	0.336	0.470 (0.470)	0.239	0.287	0.345 (0.345)
F	511.08	364.16	129.60 (145.80)	418.79	280.29	140.70 (152.70)	186.52	151.60	85.80 (101.00)
# comp	7	6	5	6	5	5	5	4	5
pred. r^2	0.678	0.741	0.530	0.670	0.683	0.580	0.492	0.568	0.461
Alignment in CA II									
q^2	0.758	0.786	0.668 (0.688)	0.709	0.737	0.649 (0.663)	0.741	0.667	0.574 (0.618)
s_{PRESS}	0.598	0.552	0.701 (0.701)	0.609	0.575	0.646 (0.646)	0.343	0.383	0.431 (0.431)
r^2	0.977	0.950	0.879 (0.886)	0.959	0.911	0.822 (0.829)	0.958	0.838	0.729 (0.757)
S	0.184	0.368	0.423 (0.423)	0.230	0.334	0.460 (0.460)	0.139	0.267	0.344 (0.344)
F	633.46	330.96	164.70 (177.10)	451.19	283.13	148.80 (156.30)	380.76	179.45	86.51 (100-20)
# comp	9	4	7	7	5	5	8	4	5
pred. r^2	0.718	0.791	0.616	0.602	0.741	0.531	0.524	0.565	0.564
Alignment in CA IV									
q^2	0.673	0.735	0.634 (0.650)	0.719	0.720	0.633 (0.643)	0.627	0.636	0.534 (0.539)
s_{PRESS}	0.686	0.617	0.734 (0.734)	0.599	0.593	0.660 (0.660)	0.408	0.400	0.450 (0.450)
r^2	0.897	0.926	0.803 (0.812)	0.949	0.918	0.765 (0.771)	0.912	0.852	0.641 (0.645)
S	0.384	0.325	0.538 (0.538)	0.254	0.321	0.528 (0.528)	0.198	0.255	0.394 (0.394)
F	240.78	347.65	108.70 (115.00)	364.55	308.74	104.90 (108.50)	237.33	199.73	72.43 (73.73)
# comp	5	5	6	7	5	5	6	4	4
pred. r^2	0.705	0.755	0.470	0.721	0.728	0.421	0.528	0.609	0.110

^{a)}For calculation of DrugScore fields, the same protein as taken for the alignment was used.

^{b)}The first value refers to the PLS models using $pK_{i,tot}$, the second value in parentheses refers to $pK_{i,PLS}$ (for explanation, see text).

Table 4.3: Statistical Results of CoMFA, CoMSIA, and AFMoC^{a)} Analyses of the Selectivity Data Built from Individual Affinity Models Using Three Alignments.

	pK _i I – II			pK _i I – IV			pK _i II – IV		
	CoMFA	CoMSIA	AFMoC ^{b)}	CoMFA	CoMSIA	AFMoC ^{b)}	CoMFA	CoMSIA	AFMoC ^{b)}
Alignment in CA I									
q^2	0.687	0.728	0.646 (0.646)	0.693	0.695	0.614 (0.614)	0.544	0.596	0.432 (0.432)
s_{PRESS}	0.671	0.625	0.722 (0.722)	0.621	0.619	0.676 (0.676)	0.451	0.422	0.499 (0.499)
r^2	0.892	0.908	0.767 (0.767)	0.897	0.886	0.741 (0.741)	0.864	0.806	0.662 (0.662)
S	0.395	0.363	0.585 (0.585)	0.359	0.379	0.554 (0.554)	0.246	0.293	0.385 (0.385)
F	221.99	267.81	85.79 (85.79)	236.06	209.83	91.58 (91.58)	172.34	141.28	50.87 (50.87)
# comp	5	5	6	5	5	5	5	4	6
pred. r^2	0.571	0.728	0.489	0.544	0.467	0.453	0.445	0.570	0.462
Alignment in CA II									
q^2	0.731	0.775	0.678 (0.678)	0.702	0.738	0.639 (0.639)	0.702	0.641	0.495 (0.495)
s_{PRESS}	0.624	0.569	0.684 (0.684)	0.616	0.574	0.654 (0.654)	0.367	0.400	0.469 (0.469)
r^2	0.935	0.927	0.786 (0.786)	0.944	0.911	0.734 (0.734)	0.886	0.869	0.701 (0.701)
S	0.307	0.325	0.580 (0.580)	0.267	0.334	0.562 (0.562)	0.227	0.241	0.361 (0.361)
F	316.05	340.39	132.60 (132.60)	320.23	277.46	88.35 (88.35)	148.20	179.27	74.91 (74.91)
# comp	6	5	4	7	5	5	7	5	5
pred. r^2	0.756	0.771	0.533	0.711	0.651	0.583	0.600	0.480	0.405
Alignment in CA IV									
q^2	0.664	0.728	0.656 (0.656)	0.697	0.704	0.655 (0.655)	0.589	0.584	0.478 (0.478)
s_{PRESS}	0.695	0.625	0.707 (0.707)	0.618	0.610	0.638 (0.638)	0.428	0.429	0.476 (0.476)
r^2	0.898	0.910	0.733 (0.733)	0.897	0.896	0.730 (0.730)	0.853	0.824	0.596 (0.596)
S	0.382	0.359	0.622 (0.622)	0.360	0.361	0.564 (0.564)	0.256	0.279	0.419 (0.419)
F	283.94	273.30	109.81 (109.81)	234.92	233.03	108.14 (108.14)	156.80	159.05	58.95 (58.95)
# comp	5	5	4	5	5	4	5	4	4
pred. r^2	0.610	0.765	0.554	0.559	0.725	0.444	0.532	0.550	0.487

^{a)} For calculation of DrugScore fields, the same protein as taken for the alignment was used.

^{b)} The first value refers to the PLS models using $pK_{i,tot}$, the second value in parentheses refers to $pK_{i,PLS}$ (for explanation, see text).

Table 4.4: Cumulative Variances^{a)} Captured in the First Three Principal Components of the GRID/CPCA Analysis.

PC	Consensus ^{b)}	Probe			
		DRY	C3	N1	O
1	55.31	42.38	60.87	61.47	59.73
2	63.80	47.75	69.11	70.04	69.10
3	67.67	55.30	72.99	73.25	71.91

^{a)}Percentage of overall variance of original descriptor matrix.

^{b)}Corresponds to a simple PCA using all probes simultaneously.

5 Untersuchung der Eignung von 3D-QSAR-Modellen für Datenbankscreening

5.1 Einleitende Bemerkung

Das folgende Kapitel wurde zum Zeitpunkt der Anfertigung dieser Dissertation als Artikel zur Veröffentlichung in der Zeitschrift *Journal of Chemical Information and Modeling* eingereicht und angenommen. Der Text erscheint daher im Folgenden in englischer Sprache.

5.2 Abstract

The applicability and scope of 3D-QSAR methods (CoMFA, CoMSIA) to screen databases is examined. A protocol requiring minimal user intervention has been established to align training and test set molecules using FlexS. As model system isozymes of human carbonic anhydrase (hCA) are used, and all results are exemplified studying affinity towards hCA II and selectivity between hCA I and II. The predictive power of the obtained models is assessed through prediction of 663 compounds not included in the training set and compared to 1D- and 2D-QSAR models derived from fragment (MACCS) or property (VSA) based descriptors. The predictive power is evaluated with respect to the following criteria: the numerical, concerning the absolute accuracy of prediction, and the categorical, characterizing the ability to assign a compound to the correct activity class.

5.3 Introduction

Quantitative Structure Activity Relationships (QSAR) describe the dependence of the biological activity or any other property under consideration (then sometimes termed

QSPR) of a set of molecules with respect to their chemical structures. The various approaches differ in the type of descriptors employed to encode the structures (e.g. an overall property, topological indices, molecular field descriptors) and in the statistical learning method used to extract a correlation (e.g. linear regression, PLS, decision trees, artificial neural networks, support vector machines). Since the introduction of Hansch [Hansch & Fujita, 1964] and Free-Wilson [Free & Wilson, 1964] analysis in 1964, QSAR techniques have become widely used and established tools in the lead optimization process [Kubinyi, 1993b].

QSAR methods require a training set of molecules for which the target property is known in advance to generate a model (except e.g. kNN-QSAR [Zheng & Tropsha, 2000] which is purely instance-based thus requiring no model deduction). Such a model can subsequently be used to predict the numeric (i.e. pK_i , IC_{50}) or categorical (i.e. active/inactive, high/low pK_i) activity of a novel molecule not contained in the training set. Furthermore, several of the statistical methods available provide information about the relevance of the applied descriptors for the activity/property being studied. For example, regression techniques return weights for each independent variable thus indicating through sign and magnitude how strongly and in which direction the respective descriptor modulates the dependent variable. This knowledge provides insights into the mechanism of action of the studied molecules and supports the synthesis planning of prospective modifications in order to improve the desired properties (direct design). However, all QSAR approaches, even those which do not produce easily interpretable output (like ANNs), can be used to predict the activity of molecules (indirect design). Thus, QSAR models can be used for lead optimization [Brown & Lewis, 2006] or database screening [Shen et al., 2004; Oloff et al., 2005; Tropsha, 2004].

Despite such potential only a limited number of studies has been communicated that actually use 3D-QSAR models for database screening. Moro et al. employed autocorrelation vectors of the molecular electrostatic potential (MEP) in conjunction with PLS analysis to screen a small-sized focused virtual combinatorial library containing 841 entries for potent antagonists of the human A_3 receptor [Moro et al., 2006]. The autocorrelation approach produces descriptors which encode the three-dimensional distribution of molecular properties, but their comparative evaluation does not depend on an alignment. Similarly, Pastor et al. use a modified auto- and cross-correlation transform to derive alignment-independent 3D descriptors from the well-known GRID molecular interaction fields. This GRIND (GRid-INdependent Descriptors) [Pastor et al., 2000] approach has been used for virtual screening [Carosati et al., 2007] and 3D-QSAR

analyses [Benedetti et al., 2004]. Murcia and Ortiz [Murcia & Ortiz, 2004] described the development of a fully automatic virtual screening workflow consisting of docking and subsequent COMBINE analysis. They established the models and simulated virtual screening experiments using factor Xa inhibitors. COMparative BINDing Energy (COMBINE) analysis [Ortiz et al., 1995] is a receptor-based 3D-QSAR approach using protein-ligand interaction energy terms as independent variables thus requiring knowledge or reasonable assumptions about the binding geometries. Zhang et al. performed a virtual screen for HMG-CoA reductase inhibitors using a combination of pharmacophore filtering, docking, and CoMFA predictions. Therein, the docking procedure provided the molecular alignment rule [Zhang et al., 2007a].

In the present study we assess the suitability of 3D-QSAR models (CoMFA [Cramer, 1988; Cramer et al., 1993] and CoMSIA [Klebe et al., 1994; Klebe, 1998]) for large scale applications. 3D-QSAR techniques require an appropriate spatial superimposition of the molecules under consideration prior to calculation of field descriptors. Various procedures have been pursued for this crucial step [Klebe, 1993; Lemmen & Lengauer, 2000] many of them being entirely manual (RMSD fit, minimization in the binding pocket of the target protein) or at least requiring user intervention (selection of docking poses). Besides the subjective nature of a user-supervised alignment, the enormous time effort makes such an approach inappropriate for large screening scenarios. Thus, we decided to develop a protocol which employs the property- and interaction-based FlexS [Lemmen et al., 1998] program as alignment engine. FlexS is fast enough for screening purposes, easily extensible to combinatorial libraries and uses a physicochemical description of the molecules in terms of Gaussian functions which is consistent with the methodology applied in CoMSIA. The statistical performance of the models obtained by this fully-automated alignment procedure is compared to the results of a study previously conducted in our laboratory [Hillebrecht et al., 2006]. There, the same dataset was used, but the molecules were manually aligned and minimized in the binding pocket.

In addition to 3D-QSAR we also derived 1D and 2D models and applied both to an external test set of 663 molecules. Two kinds of descriptors are used: The public version of the well-known MDL MACCS keys [McGregor & Pallai, 1997; Durant et al., 2002], which encode the presence or absence of distinct structural fragments in a molecule, and the 32 partitioned van-der-Waals surface area (VSA) descriptors developed by Labute [Labute, 2000], which are property based. The performance in terms of external predictivity is compared on numerical and categorical basis: predictive r^2 and Spearman's r^2

assess the correlation between experimental and calculated activity values, Spearman's rank correlation coefficient characterizing the ability to correctly predict the rank order of the compounds according to their activity. While these parameters characterize the overall numerical accuracy of the models, the ability to enrich compounds with a desired activity is even more important for database screening purposes. In contrast to other virtual screening approaches such as docking/scoring or pharmacophore matching, however, one limitation of QSAR is that it works only on a rather limited piece of chemical space covered by the training set compounds. Hence, only focused libraries can be screened reliably, and in order to obtain enrichment or classification ratings a certain activity threshold has to be predefined for class assignment (e.g. if $\text{pK}_i > 7$, compound is classified "active"). In this study, we report sensitivity and specificity, classical enrichment plots, receiver operating characteristic (ROC) curves, the area under the ROC curve (AUC) and hit rates to compare different models from an objective point of view [Witten & Frank, 2005; Triballeau et al., 2006].

All studies reported herein are based on pK_i values of sulfonamide-type inhibitors towards human carbonic anhydrase (hCA) isozymes. CAs are zinc containing hydrolases (EC 4.2.1.1), which catalyze the reversible hydration of carbon dioxide to bicarbonate under release of one proton [Maren, 1967; Lindskog, 1997]. They are involved in a variety of important physiological processes, such as pH and CO_2 regulation, bone resorption, calcification, metabolic reactions, tumorigenicity, and electrolyte secretion [Geers & Gros, 2000; Supuran & Scozzafava, 2001]. Therefore, inhibitors of carbonic anhydrases offer the opportunity to treat several physiological disorders, e.g. as drugs against glaucoma, mountain sickness, and epilepsy [Supuran & Scozzafava, 2000b]. Due to the high structural similarity and the broad tissue distribution of the different isoforms of CAs selectivity is an issue of major concern in the design of CA inhibitors. In the present study we used the pK_i value measured against hCA II (in the following sections referred to as $\text{pK}_i(\text{II})$) as an example for affinity prediction and the difference of the pK_i values towards hCA I and hCA II (in the following sections referred to as $\Delta\text{pK}_i(\text{I} - \text{II})$) as an example for selectivity prediction.

5.4 Methods

All 3D-QSAR studies and all PLS analyses were performed using SYBYL7.1 [SYBYL 7.1]. MACCS keys [McGregor & Pallai, 1997; Durant et al., 2002] and VSA descriptors

[Labute, 2000] were calculated using MOE 2005.06 [MOE 2005.06]. VSA descriptor selection was accomplished by the SVL script (Scientific Vector Language) AutoQSAR [AutoQSAR]. Automation of ligand alignment was realized via the Python interface “pyflexs” running FlexS 1.20.2 [Lemmen et al., 1998]. Compound prediction was implemented in SPL (Sybyl Programming Language).

5.4.1 Dataset and Preparation

In order to obtain a suitable training set 144 ligands were taken from our previous 3D-QSAR study and subjected to the FlexS alignment procedure. For details about selection criteria and composition of this dataset we refer to reference [Hillebrecht et al., 2006]. The following procedure was pursued to automate ligand alignment: For each of the nine scaffolds (see Fig. 4.1 in Chapter 4.2, throughout this paper we will refer to these scaffolds by the corresponding capital letters **A** to **I**) comprised in training and test set two items had to be defined: (1) a so-called MAPREF which is an anchoring fragment used to initiate the incremental construction algorithm of FlexS and (2) a reference ligand for spatial alignment. The assignment of atoms in each ligand to the corresponding MAPREF substructure is achieved by graph matching and does not require any user intervention. The second item to be predefined is one complete reference ligand onto which the respective candidate molecules are superimposed. This reference should be at best a compound with maximum spatial extensions adopting a “representative” conformation for the individual chemical class, if possible extracted from a crystal structure. The utilization of a MAPREF substructure as alignment template ensures internal consistence of the produced alignment. Without such constraints placements lacking proper superimposition of the anchoring fragments with that of the reference ligand might achieve higher similarity scores because the overall similarity comprising various properties of the entire molecule of a distinct pose can be higher than the score obtained for the desired pose where common substructures are mutually overlaid. Using this procedure all except six molecules could be aligned ending up with a final training set size of 138 molecules.

The training set protocol was applied to 663 compounds serving as a real life test example. It has to be noted that this test set represents a rather unbalanced sample since the only selection criteria were availability of experimental data to be predicted and affiliation to one of the crude chemotypes considered in the training set. Most

of the compounds could be aligned applying default settings in FlexS, only for some rather rigid and large molecules the threshold for the minimum van-der-Waals overlap volume had to be reduced from 0.6 to 0.4 in order to obtain a superimposition solution.

5.4.2 3D-QSAR Analyses

For CoMFA [Cramer, 1988] analysis the interaction energies between a probe atom and the ligand atoms were calculated using a grid box of $26 \times 34 \times 25$ points with 1 Å spacing, embedding all ligands with a margin of at least 4 Å in each direction. The same box dimensions were also used for CoMSIA [Klebe et al., 1994] studies. A positively charged sp^3 -carbon atom was used as probe atom for calculating steric and electrostatic CoMFA fields applying SYBYL standard parameters (TRIPOS standard field, dielectric constant $1/r$, cutoff 30 kcal/mol). CoMSIA fields were computed for steric, electrostatic, hydrophobic, and hydrogen bonding properties, using a probe of charge +1, radius of 1, hydrophobicity and hydrogen bonding properties of +1, and an attenuation factor α of 0.3 for the Gaussian distance-dependent function. All fields were scaled with the CoMFA_STD scaling procedure, assigning equal weights to each field. The response variables ($pK_i(\text{II})$, $\Delta pK_i(\text{I} - \text{II})$) were correlated with the field descriptors using SAMPLS [Bush & Nachbar, 1993] in a leave-one-out cross-validation analysis. The optimal number of PLS components was determined by subsequently extracting one more latent variable until the corresponding q^2 value is not further increased by more than five percent [Thibaut et al., 1993]. Afterwards, a PLS analysis [Wold et al., 1993] was performed without cross-validation using the optimal number of components, applying no column filtering.

5.4.3 1D- and 2D-QSAR Analyses

The MDL MACCS keys have already proved useful for screening purposes in conjunction with PLS Discriminant Analysis (PLS-DA) models [Evers et al., 2005]. We decided to use them for building numeric 2D-QSAR models. The public version was used that evaluates the presence of 166 distinct molecular fragments. Instead of computing binary fingerprints which store the information about presence or absence of one particular substructure the “counted” version was used resulting in 166 integers each capturing the exact number of occurrences of the fragment in a molecule. This

protocol allows better differentiating molecules possessing not only qualitatively the same substructures but differing in counts. This is often the case in focused libraries.

In addition to such fragment oriented models also the property based VSA descriptors were used to derive 1D-QSAR models. Each of the 32 VSA descriptors is calculated as the sum of all atomic contributions to the approximated van-der-Waals surface areas with the respective property falling into a certain range. In MOE this type of descriptor can be calculated for three properties: SlogP (calculated octanol/water partition coefficient), SMR (calculated molar refractivity), and PEOE (Gasteiger-Marsili partial charges). These descriptors show a rather low degree of correlation among each other, capture different aspects of protein-ligand interactions and transport phenomena, and their broad applicability to many QSAR related problems has been shown. In order to select a suitable subset of descriptors the AutoQSAR [AutoQSAR] procedure was applied. It identifies the “best” model based on leave-one-out cross-validation. This procedure should only generate rather crude 1D-QSAR models. It will not evaluate the capacity of VSA descriptors exhaustively, and the obtained models should rather serve as a first overview for comparison to get an idea what can be achieved by a crude “quick and dirty” 1D model.

5.4.4 Evaluation of Predictivity

In order to assess and compare the predictive power of the different models, several statistical parameters as well as plots are reported: The predictive r^2 value is usually used to characterize the performance in terms of external predictivity of QSAR models. Additionally, Spearman’s rank correlation coefficient was calculated which quantifies the ability of a model to correctly predict the relative order instead of the absolute numeric value of the modeled variable.

Since our study is intended to assess the predictive power of QSAR models for database screening purposes it is reasonable to apply also figures-of-merit commonly used in virtual screening, where the correct prediction of class memberships is even more important. Therefore, a somewhat arbitrary threshold had to be defined which determines class affiliation: For $pK_i(II)$ the 2% or 5% of molecules with the highest (or lowest, respectively) activity were considered as the “high activity” (or “low activity”) class, whereas the remaining part of the molecules is assigned to the complementary class. This setting simulates screening experiments where one wants to enrich compounds

which possess a remarkably high affinity towards the target of interest, compared to the bulk of the training set. Defining the 2% or 5% of lowest activity as the class of interest, the situation represents antitarget modeling, where a distinct receptor must not be inhibited (this holds particularly for e.g. Cytochrome P450 or hERG channel blockers). The same aspects are examined for the activity difference $\Delta pK_i(I - II)$. This corresponds to a screening scenario with the aim of enriching compounds selective towards hCA I or hCA II, respectively. As direct measures for correct class prediction the sensitivity Se (also called “true positives rate” or “recall”), the specificity Sp , and the hit rate H (or “precision”) are reported [Witten & Frank, 2005; Triballeau et al., 2006]:

$$Se = \frac{TP}{TP + FN} \cdot 100\%$$

$$Sp = \frac{TN}{TN + FP} \cdot 100\%$$

$$H = \frac{TP}{TP + FP} \cdot 100\%$$

where FP and FN is the number of false positives and false negatives, and TP and TN is the number of true positives and negatives, respectively.

To give an even more comprehensive illustration of the results several kinds of plots are reported:

- A plot displaying the predicted activity/selectivity value of a molecule on the ordinate versus its experimental one on the abscissa.
- A classical **enrichment plot**. On the x-axis this graph displays the amount of database entries screened, on the y-axis it shows the amount of actives retrieved from the database.
- A **receiver operating characteristic (ROC)** curve. This type of plot is quite popular in many scientific areas like psychology, medicine, acoustics, or criminology to assess the ability of a diagnostic system to distinguish signal from noise. On the x-axis a ROC plot displays the term $1 - \text{specificity}$ which corresponds to the “noise” in the dataset, on the y-axis it shows the sensitivity which can

be thought of as the “signal” that is to be identified by the ranking procedure. Nevertheless, its application in drug design and especially in virtual screening is still not standard although it exhibits some advantages compared to the usually applied enrichment curves [Triballeau et al., 2006, 2005]. Most important the **area under the curve** (AUC) of a ROC plot can be used to directly compare the achieved accuracy of a computer test. Furthermore, the shape of the “ideal curve” of an enrichment plot depends on the ratio of actives to inactives in the database, which is not the case for the ROC curves. Finally, enrichment plots only capture one aspect of a screening experiment, namely the power to retrieve actives, i.e. the sensitivity, whereas ROC curves also illustrate the second important aspect, the ability to discard inactives, i.e. the specificity.

5.5 Results and Discussion

5.5.1 Comparison of 3D- and 2D-QSAR Models in Terms of Internal Predictivity

To get an impression about the internal consistency, Figure 5.1 displays the leave-one-out q^2 value as a crude measure of model performance. For all methods q^2 is smaller for selectivity prediction ($\Delta pK_i(I - II)$) than for affinity prediction ($pK_i(II)$) which is a consequence of the fact that $\Delta pK_i(I - II)$ is a difference of two single variables and thus contains the sum of both errors. In the case of CoMFA and CoMSIA the automated alignment yields slightly smaller q^2 values (largest difference between manual and automated alignment: 0.055 for CoMFA, $pK_i(II)$). The observation that the differences with respect to the alignment method are smaller for CoMSIA can be attributed to the smoother Gaussian functional form used to derive the descriptors which makes this method less sensitive towards slight shifts of the molecules. Since the overall performance of the 3D-QSAR methods is virtually the same it can be stated that the automated alignment procedure is suitable for large scale applications.

The 1D- and 2D-QSAR approaches produce models of similar quality for prediction of $pK_i(II)$, however, in case of selectivity prediction the decrease in q^2 is significantly higher compared to the 3D methods. This decrease is more pronounced for the VSA descriptors compared to the MACCS keys. Nevertheless, still highly significant models

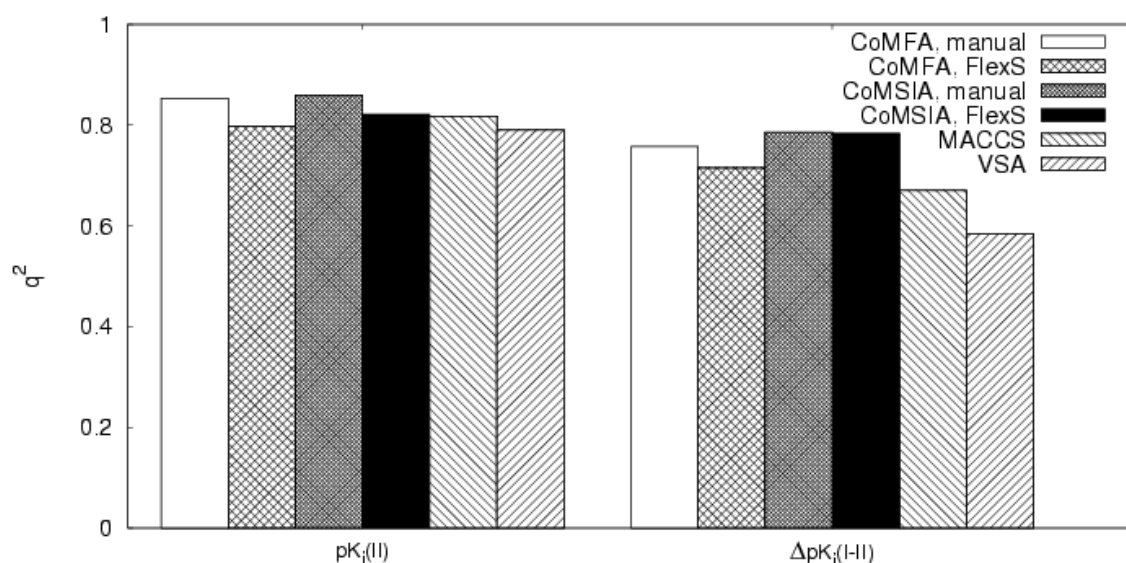


Figure 5.1: Comparison of performance of different models evaluated via cross-validation. The leave-one-out- q^2 is shown for prediction of $pK_i(II)$ (left) and $\Delta pK_i(I - II)$ (right), respectively.

($q^2 > 0.5$) can be obtained.

The full set of statistical parameters is shown in Table 5.1.

5.5.2 External Numerical Predictivity: Correlation Coefficient

A test set of 663 sulfonamide type inhibitors was aligned with the above-mentioned FlexS protocol, subsequently $pK_i(II)$ and $\Delta pK_i(I - II)$ were predicted based on the 3D-QSAR models previously obtained by FlexS alignment and the 1D/2D models (independent on conformation and orientation). Table 5.2 shows the predictive r^2 and Spearman's rank correlation coefficient for the different models. CoMSIA yields the best results in all cases, CoMFA exhibiting slightly worse values. The MACCS approach performs significantly better than VSA, however, its predictive power is clearly worse compared to the 3D techniques. Regarding the size and structural imbalance of the test set with respect to the training set the results for prediction of $pK_i(II)$ are remarkably good for the 3D methods and still acceptable for the MACCS approach, whereas the VSA approach fails to make useful predictions. These findings confirm the well-known “beware of q^2 !” phenomenon, stating that a “good q^2 ” is by no means indicative that a model also possesses sufficient predictivity with respect to novel external compounds

Table 5.1: Statistical Results of the Different QSAR Analyses.

	CoMFA		CoMSIA		MACCS	VSA
alignment	manual	FlexS	manual	FlexS	—	—
pK _i (II)						
q ²	0.853	0.798	0.860	0.822	0.818	0.790
SPRESS	0.504	0.593	0.489	0.552	0.556	0.629
r ²	0.949	0.880	0.943	0.867	0.884	0.837
S	0.297	0.457	0.313	0.476	0.444	0.553
F	423.84	243.85	453.84	441.59	264.53	27.33
# comp.	6	4	5	2	4	17
Δ pK _i (I – II)						
q ²	0.758	0.715	0.786	0.784	0.670	0.584
SPRESS	0.598	0.629	0.552	0.548	0.684	0.799
r ²	0.977	0.851	0.950	0.905	0.802	0.660
S	0.184	0.455	0.368	0.364	0.529	0.720
F	633.46	190.35	330.96	316.40	188.96	14.03
# comp.	9	4	4	4	3	13

Table 5.2: Numerical Measures of Predictivity for Different QSAR methods.

		CoMFA	CoMSIA	MACCS	VSA
pred. r ²	pK _i (II)	0.454	0.482	0.302	-0.710
	Δ pK _i (I – II)	-0.079	0.001	-0.476	-0.893
Spearman's r ²	pK _i (II)	0.407	0.443	0.393	0.288
	Δ pK _i (I – II)	0.115	0.118	0.069	0.000

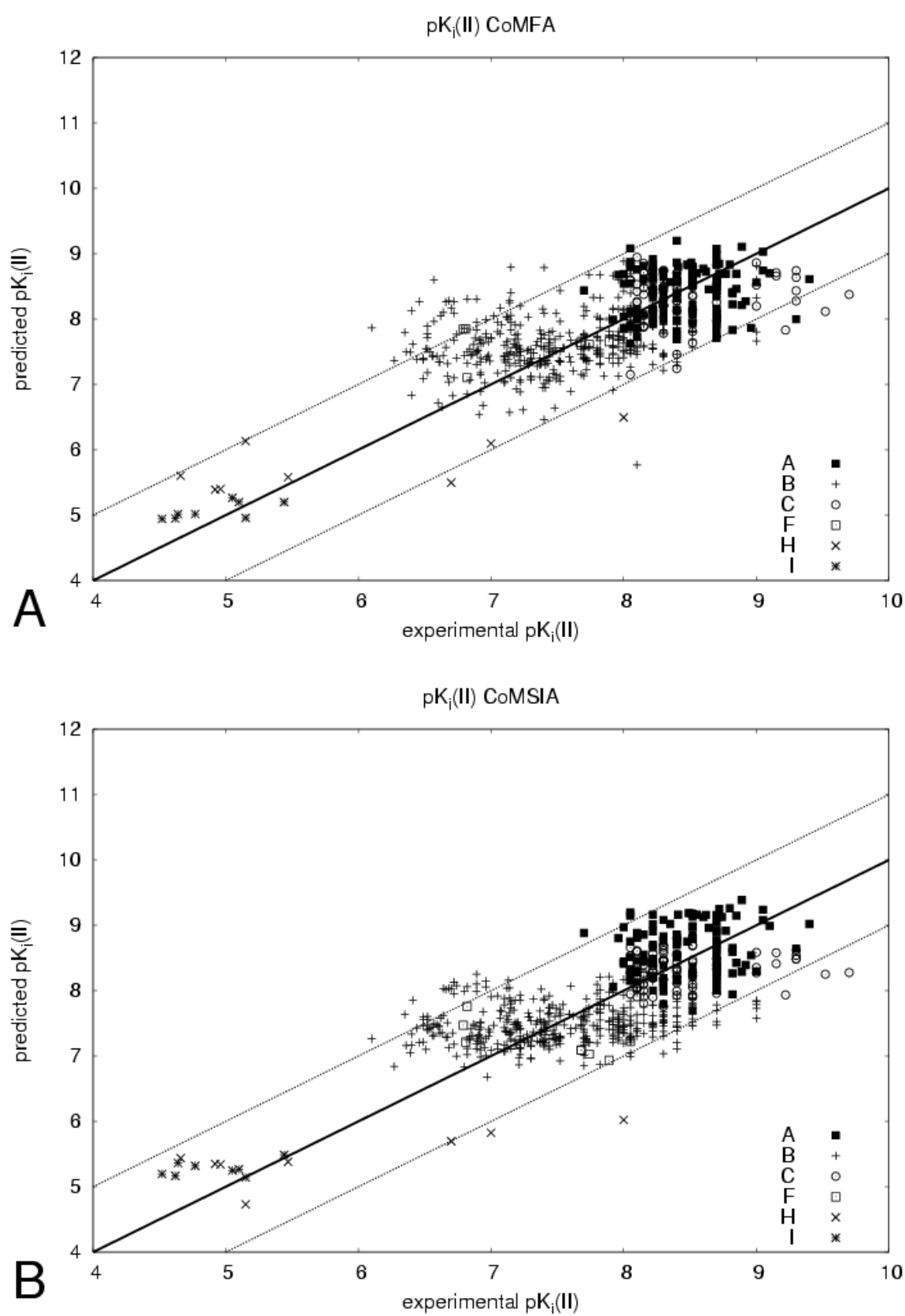
not regarded for training. Though, the relative ranking of the model quality can still be estimated based on the achieved q² in our study. Obviously all approaches are not able to give accurate numerical predictions for the selectivity variable Δ pK_i(I – II). Nevertheless, we will show in the following sections that these models are not completely useless with respect to screening purposes for which accurate numerical predictions are less important than correct assignment to affinity classes.

5.5.3 External Numerical Predictivity: Experimental versus Predicted Plots

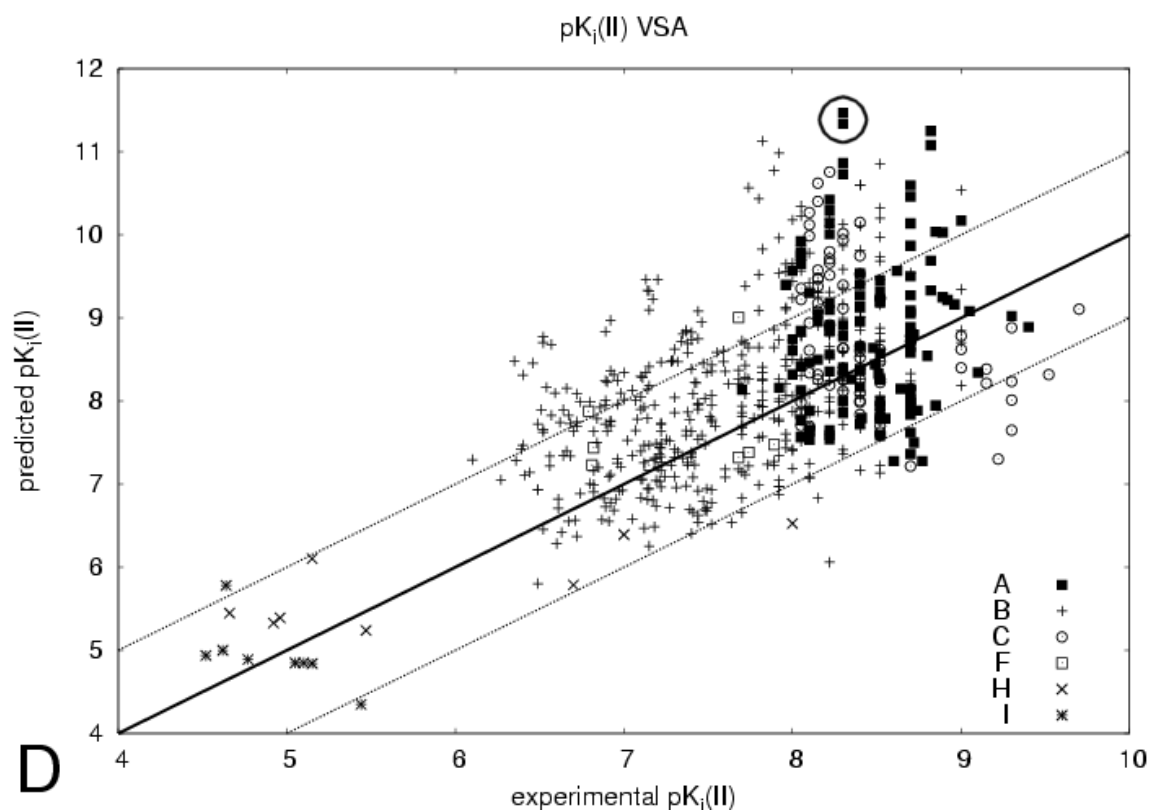
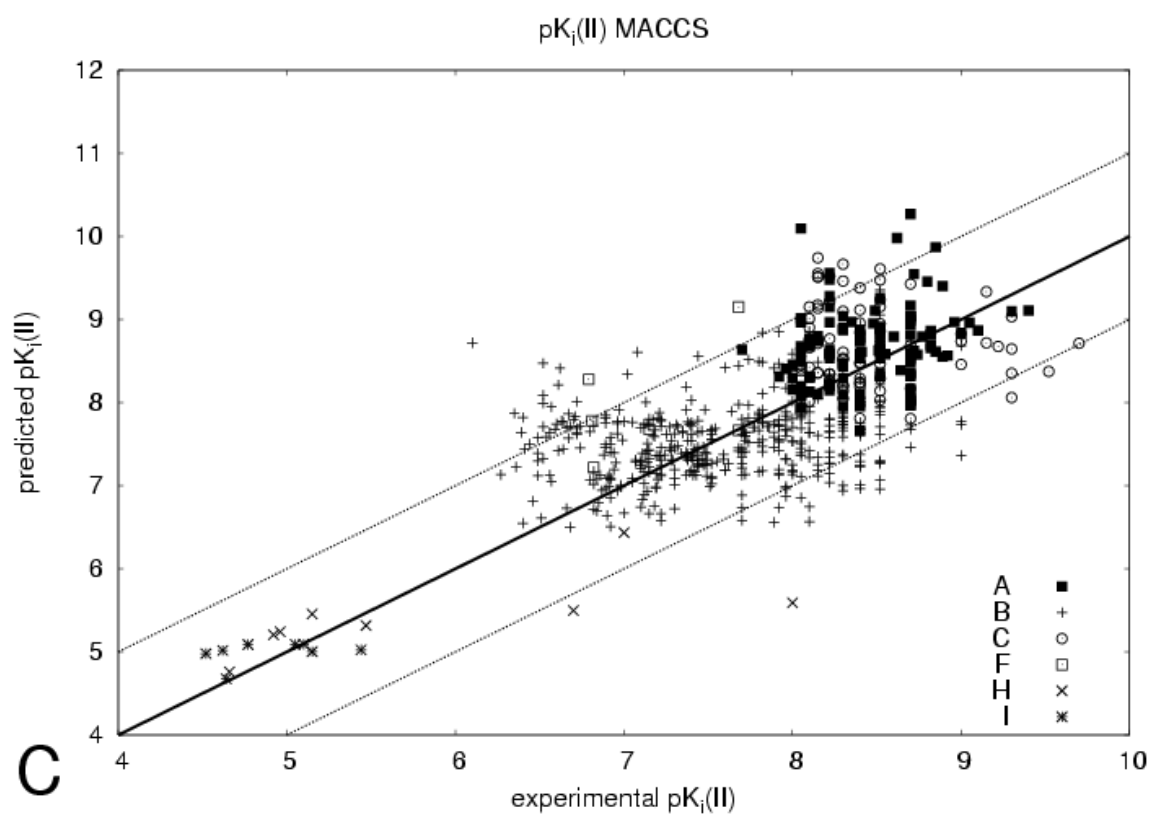
Figure 5.2 shows the plots of experimental versus predicted $pK_i(\text{II})$ values for all four approaches. Each point corresponds to one molecule and indicates its membership to a chemical class by point shape. The distribution of points along the abscissa (i.e. the experimental $pK_i(\text{II})$) reflects again the imbalance of the test set. Compounds with $pK_i(\text{II})$ values below 6.5 are rarely found, whereas those possessing $pK_i(\text{II})$ between 6.5 and 9.0 are clearly overrepresented. The plots demonstrate the value of CoMFA (Figure 5.2(a)) and CoMSIA (Figure 5.2(b)) for this large scale application since most of the test compounds are predicted correctly within ± 1 logarithmic unit. The same holds for the MACCS key approach (Figure 5.2(c)) albeit with some more molecules falling outside this tolerance. Closer inspection of the plots reveals a tendency of the MACCS keys and even more of the VSA descriptors (Fig. 5.2(d)) to overpredict some of the thiadiazolsulfonamides (scaffold **A**, Fig. 4.1) and the benzothiazolsulfonamides (**C**). In order to shed some light on the reason for this finding, we pick two molecules marked by a bold circle in Figure 5.2(d) whose affinity is predicted more than 3 logarithmic units too high by the VSA model whereas the other approaches provide a reasonable estimate. A projection of these molecules into PCA space of the training set calculated using VSA descriptors reveals them as extreme structural outliers. They contain very bulky phenylpyridinium groups which are not present in any of the training set compounds. The most relevant descriptor responsible for the overprediction is SMR_VSA5 (strongly correlated to molecule volume and polarizability) which for most of the training set molecules does not exceed an upper limit of 160, but takes values of 212 and 231 for both ill-predicted compounds, respectively.

Figure 5.2: (see following two pages)

Plots of experimental versus predicted $pK_i(\text{II})$ values for the four QSAR approaches considered in this study. A: CoMFA; B: CoMSIA; C: MACCS; D: VSA; Dashed lines mark a range of ± 1 logarithmic unit deviation, the solid line indicates perfect correlation. The shape of the points indicate the individual scaffold class of the respective ligands the capital letters corresponding to the structural assignment given in Fig. 4.1. The bold circle in (d) marks two extreme outliers discussed in the text.



(figure continued on next page)

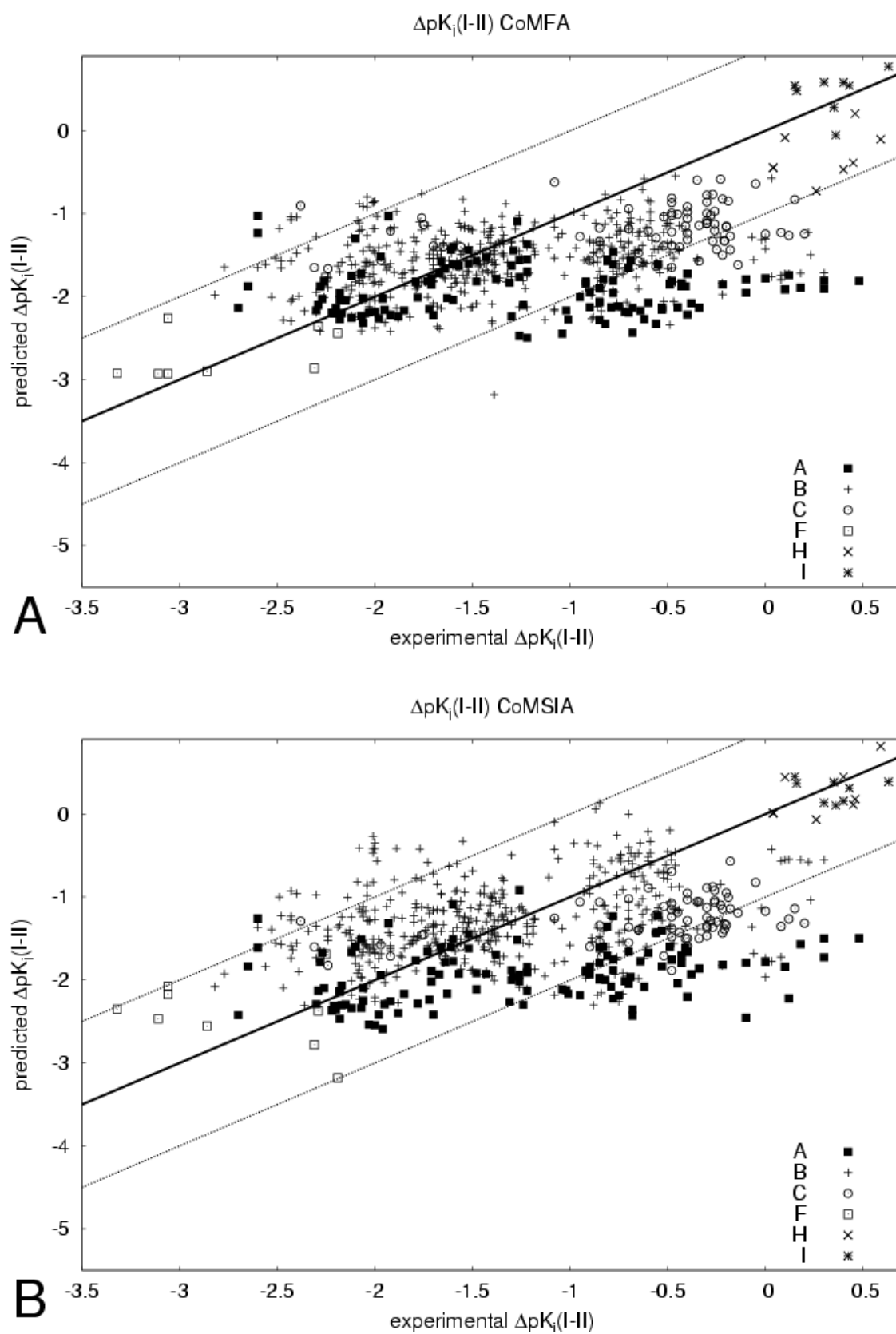


Obviously, the field- and fragment-based descriptors are more robust with respect to structural extrapolation, at least in the case of our data set. It has to be noted, however, that a descriptor used to evaluate a compound in CoMFA/CoMSIA contains several thousand values, however, for MACCS only 166, and for VSA even only 32, respectively. Thus, it is not too surprising that a set of VSA values is only a crude approximative description of a molecule. Moreover, no information about the spatial distribution of the properties encoded by the VSA descriptors about the molecules is contained in the latter descriptors.

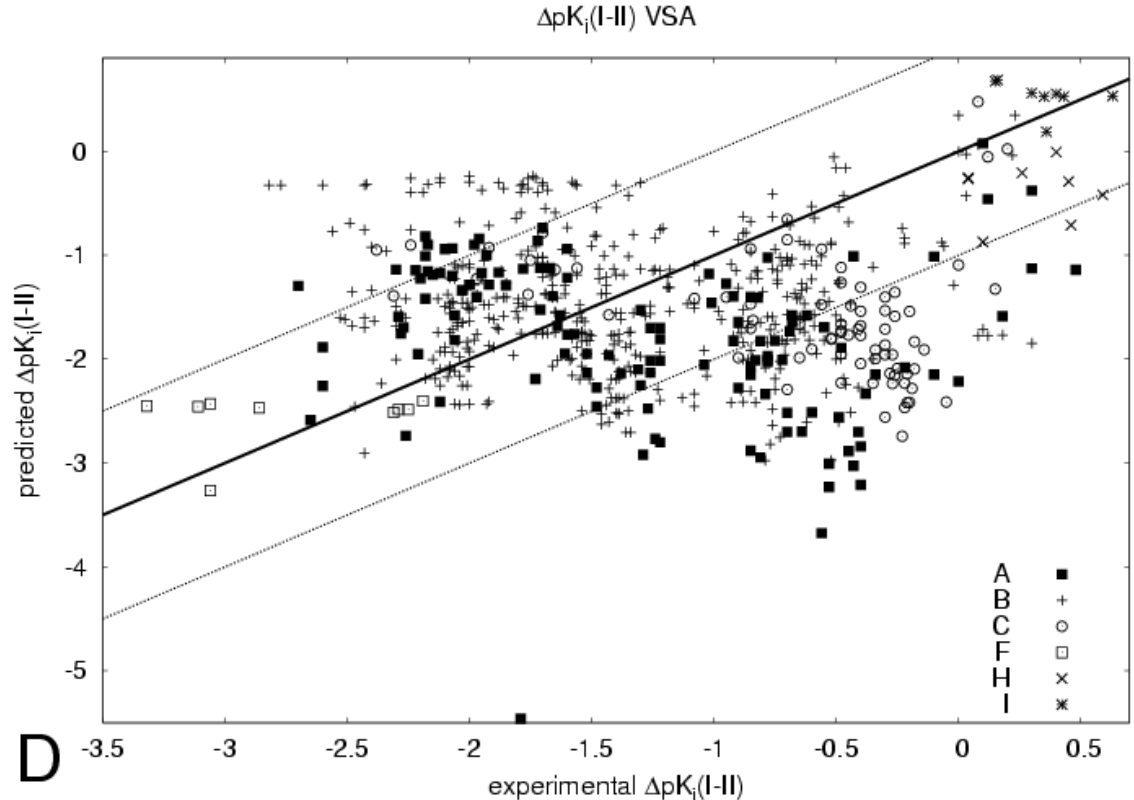
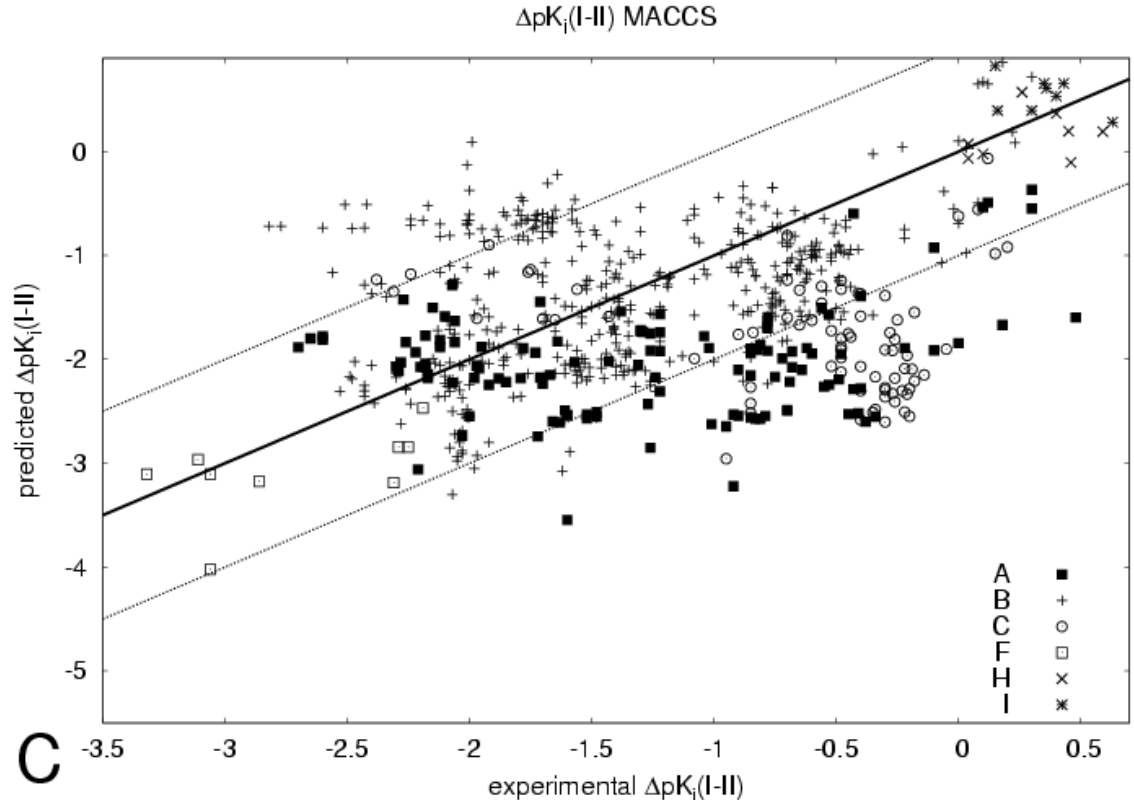
As expected deviations from correct predictions are generally worse for selectivity prediction of $\Delta pK_i(I - II)$ as shown in Fig. 5.3. However, despite the poor predictive r^2 for all models the overall appearance of the plots suggests that the 3D-QSAR models can give rough estimates in terms of selectivity. The group of thiadiazolesulfonamides (scaffold **A**, Fig. 4.1) with experimental $\Delta pK_i(I - II) > -1$ is remarkably underpredicted by all four models. This can be easily explained by the fact that in the training set this chemical class has a mean $\Delta pK_i(I - II)$ of -1.91 with a maximum of -0.71. The poorly predicted subset, however, exhibits a mean $\Delta pK_i(I - II)$ of -0.52. Correct prediction would require clear extrapolation. In the case of benzothiazolsulfonamides (scaffold **C**) the MACCS and VSA models perform significantly worse compared to the 3D models.

Figure 5.3: (see following two pages)

Plots of experimental versus predicted $\Delta pK_i(I - II)$ for the four QSAR approaches considered in this study. A: CoMFA; B: CoMSIA; C: MACCS; D: VSA. Dashed lines mark a range of ± 1 logarithmic unit deviation, the solid line indicates perfect correlation. The shape of the points indicates the individual scaffold class of the respective ligands the capital letters corresponding to the structural assignment given in Fig. 4.1.



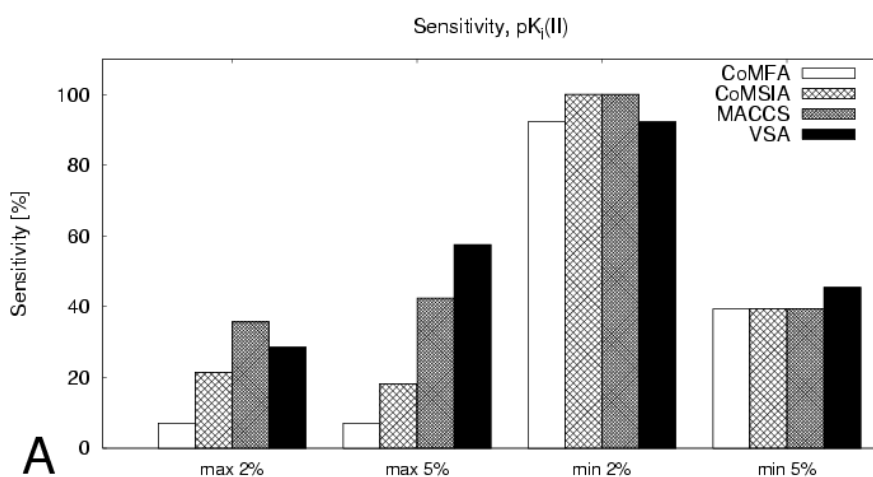
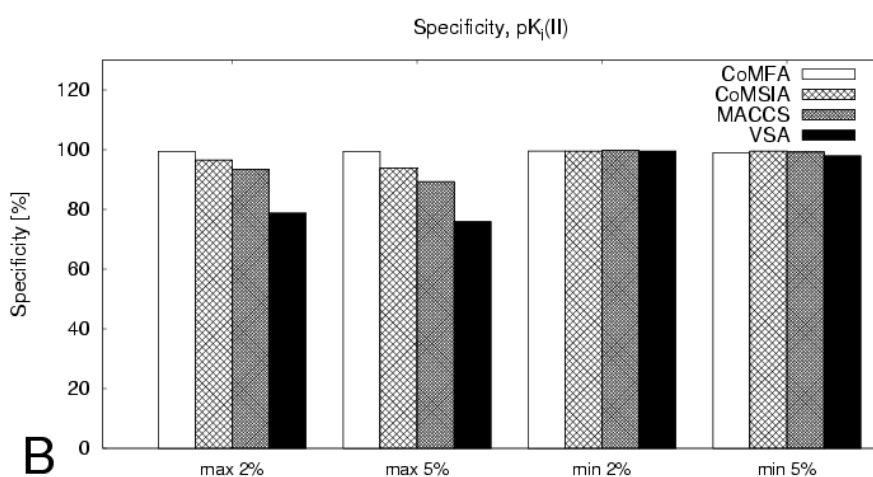
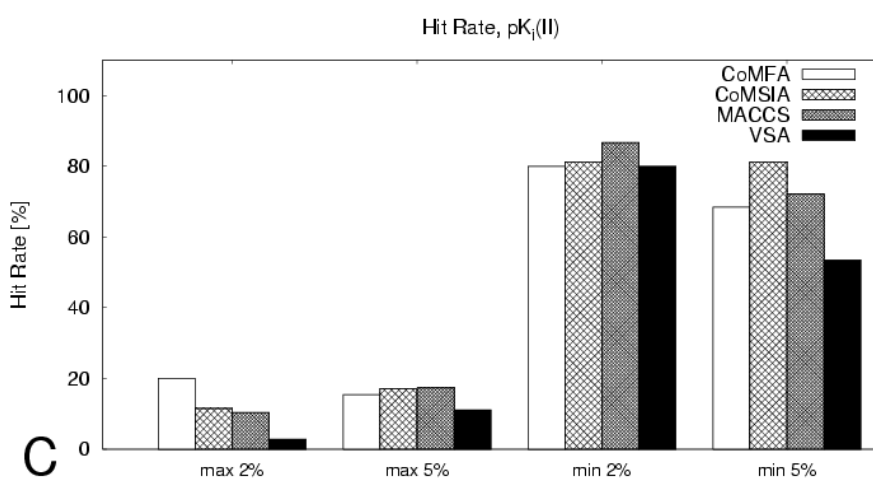
(figure continued on next page)



5.5.4 External Categorical Predictivity: Sensitivity, Specificity, Hit Rates

In order to assess the categorical external predictivity of the established models we will first report the resulting sensitivities (Se), specificities (Sp), and hit rates (H) after a distinct threshold for the modeled variable has been defined. Each compound is labeled “positive” or “negative” according to the above-mentioned arbitrarily chosen threshold. The thresholds are selected such that compounds with the highest/lowest 2% or 5% of $pK_i(\text{II})/\Delta pK_i(\text{I} - \text{II})$ are retrieved. In a screening scenario one is usually only interested in identifying compounds with “extreme” activities either to find those with high affinity for a particular target, with low affinity for an antitarget or with extraordinary selectivity profiles. In a real life situation an even lower amount ($\leq 1\%$) would be of interest due to the large size of databases screened, but for our present study this would result in a very small absolute number of molecules. Most likely rather unstable statistical results would be suggested. The same threshold will then be applied to the calculated values, and the comparison with the experimentally determined classification yields the assignment to “true positives/negatives” and “false positives/negatives” (TP, TN, FP, FN). The main difference between this kind of evaluation and the ROC and enrichment plots described below is that the latter methods monitor the evolution of Se or Sp in dependency on a variable threshold whereas the approach applied here analyses the classification performance taking the predicted values “as is”. The resulting sensitivities, specificities, and hit rates are shown in Figure 5.4 for the various approaches and thresholds. The plots reveal a trend towards the 1D/2D methods exhibiting a higher sensitivity, i.e. they tend to omit less actives than the 3D models. This is achieved at the cost of a reduced specificity as they also label many inactives as actives. This results in most cases in higher hit rates for the 3D-QSAR approaches. The MACCS keys-based models perform significantly better compared to the VSA models; they yield high sensitivities in conjunction with reasonable specificities.

The plots also show that the results are generally better to identify compounds with minimal $pK_i(\text{II})$ or maximal $\Delta pK_i(\text{I} - \text{II})$, respectively, compared to the case of maximal $pK_i(\text{II})$ or minimal $\Delta pK_i(\text{I} - \text{II})$. This finding corresponds to the observation described above that compounds with very low $pK_i(\text{II})$ (i.e. the classes of scaffolds **H** and **I**) are usually well predicted, whereas those with high $pK_i(\text{II})$ (particularly the scaffolds **A** and **C**) are often overpredicted. In consequence corresponding results are obtained for the difference $\Delta pK_i(\text{I} - \text{II})$.

(a) Sensitivity, $pK_i(\text{II})$ (b) Specificity, $pK_i(\text{II})$ (c) Hit Rate, $pK_i(\text{II})$

(figure continued on next page)

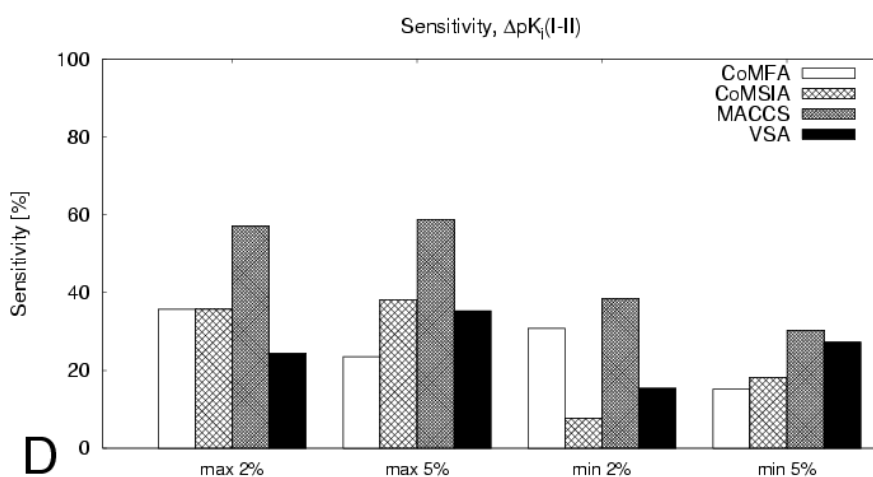
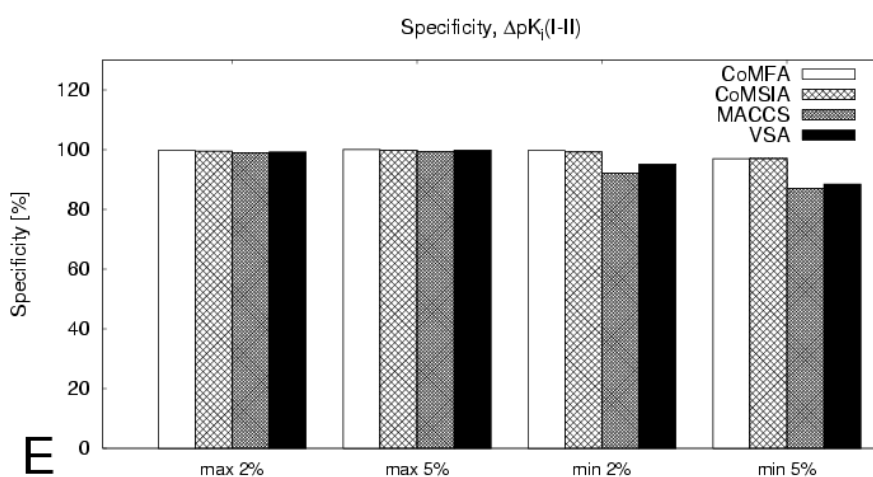
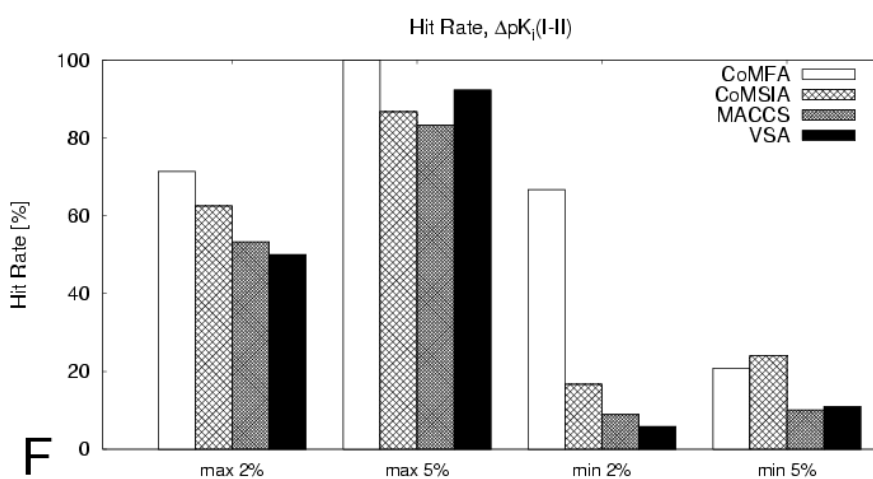
(d) Sensitivity, $\Delta pK_i(I-II)$ (e) Specificity, $\Delta pK_i(I-II)$ (f) Hit Rate, $\Delta pK_i(I-II)$

Figure 5.4: Comparison of sensitivity, specificity, and hit rate for the four QSAR methods.

5.5.5 External Categorical Predictivity: ROC Plots

Figure 5.5 shows two examples of ROC plots monitoring the screening progress. The main diagonal corresponds to a random classifier unable to discriminate signal from noise. Thus, for any possible threshold the same percentage for sensitivity (signal) and $1 - \text{specificity}$ (noise) is achieved. Its AUC is 50.0%, and any classifier better than random therefore has to produce an AUC above this lower limit. The curve of an ideal classifier coincides with the left and the top edge of the coordinate system and encloses an AUC of 100.0%.

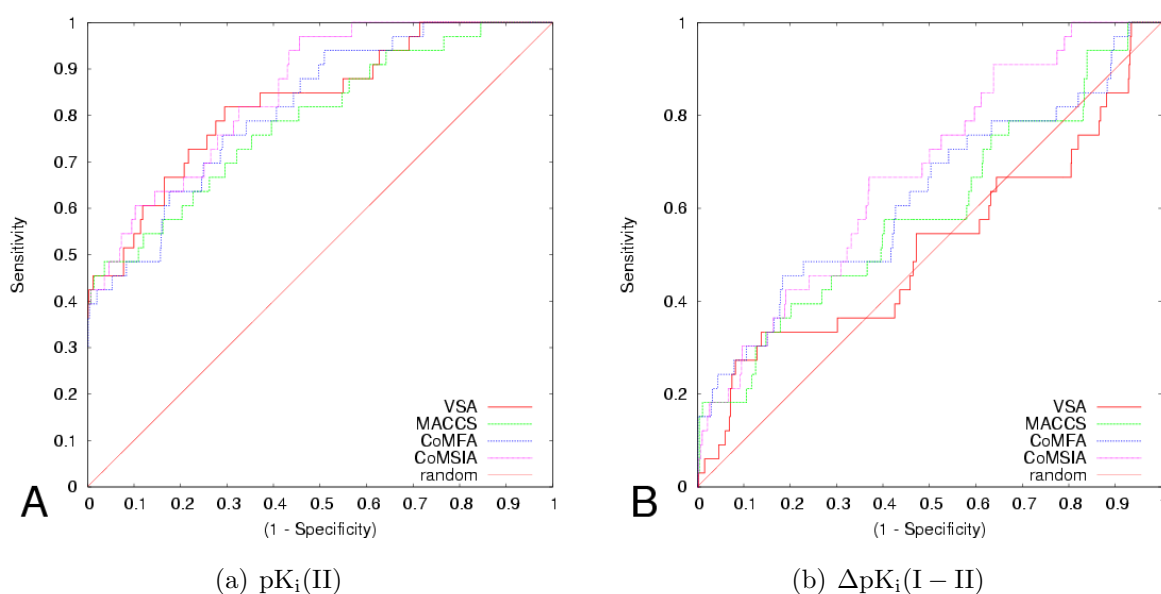


Figure 5.5: Receiver operating characteristic (ROC) curves for the retrieval of the 5% compounds with lowest $pK_i(\text{II})$ (A) and $\Delta pK_i(\text{I} - \text{II})$ (B), respectively, applying the four QSAR approaches considered in this study. The main diagonal corresponds to a random classifier whereas the left and top edge of the plot would show the ROC line of an ideal classifier.

Fig. 5.5(a) shows the ROC plots for the retrieval of the 5% compounds with lowest $pK_i(\text{II})$. For this example all four methods perform similarly well. For a real life scenario the left section of the plot is most interesting since it indicates how much signal (actives) can be identified by the model still discarding most of the noise (inactives). In our case about 35% of actives can be retrieved without extracting false positives. For the remaining part of the screening, the MACCS descriptors perform slightly worse compared to the other models. At higher noise levels ($(1 - \text{specificity}) = 0.4$) the 3D methods slightly outperform the 1D and 2D approaches.

In Fig. 5.5(b), the ROC curves to identify the 5% compounds with lowest $\Delta pK_i(I - II)$ (i.e. the most selective ones for hCA II) are shown. Here, the differences between the four approaches are more pronounced. Worst performance is indicated for the VSA model intermediately even dropping below the random line. All approaches exhibit poorer performance in the “interesting” left part of the plot compared to the previous example. In this area MACCS and 3D models perform similarly, whereas at specificities below 80% ($(1 - \text{specificity}) > 0.2$) 3D-QSAR models outperform the 1D/2D models.

In order to present a concise comparison for the other retrieval experiments, Figure 5.6 shows bar plots denoting the corresponding AUCs. For classification with respect to $pK_i(II)$ (Figure 5.6(a)), all approaches perform comparably well, however, the VSA models tend to be worse in identifying compounds with maximum $pK_i(II)$. In the case of selectivity prediction ($\Delta pK_i(I - II)$, Fig. 5.6(b)) the 1D/2D models outperform 3D-QSAR in successfully retrieving the most hCA I selective compounds, whereas the opposite is true for identification of hCA II selective compounds. As demonstrated above, 3D-QSAR methods obviously suffer less from the overprediction problem than the 1D/2D techniques. The decreasing performance of the 3D models to classify the most hCA I selective compounds results from the fact that already the individual $pK_i(I)$ values for this isoform are modeled with reduced accuracy (data not shown).

It has to be noted that the AUCs of the ROC curves allow for an overall comparison of the different approaches. However, the detailed characteristics of a classifier’s performance can only be deduced considering the overall shape of a ROC curve.

5.5.6 External Categorical Predictivity: Classical Enrichment Plots

An important information which cannot be derived from the ROC plots is the amount of actives identified when a certain subset of the database is screened. This type of information is provided by the classical enrichment plots usually used to visualize the screening performance in drug design. Similarly to the ROC curves the left part of the plots is the most interesting for practical applications. We will illustrate such an evaluation using the same examples as for the ROC plots.

Figure 5.7(a) shows the plot for the enrichment of the 5% compounds with lowest $pK_i(II)$. Within the first few percent of the screened database all four methods successfully separate actives from inactives. For the remainder of the screening process all

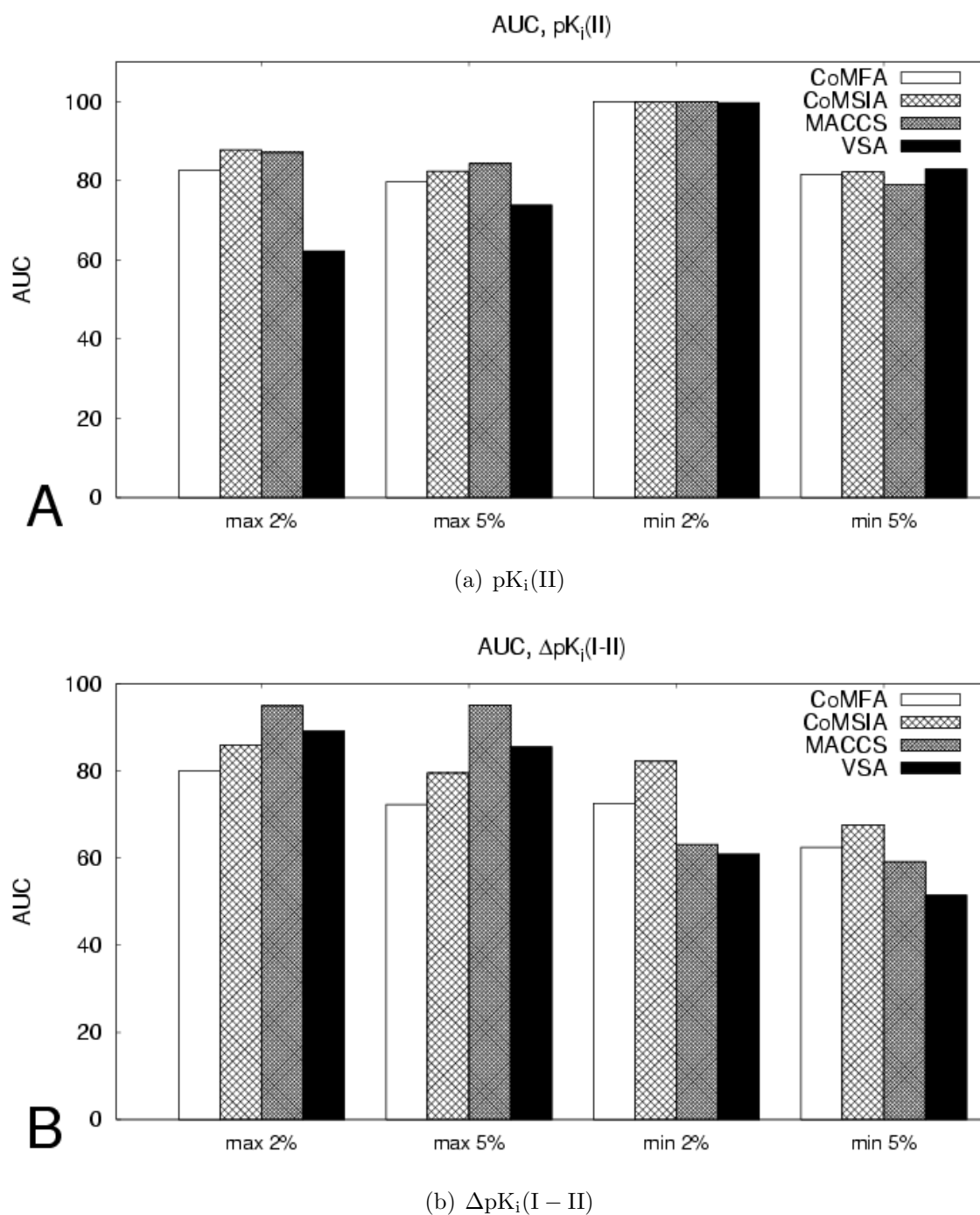


Figure 5.6: Bar plots displaying the area under the curve (AUC) values of the ROC curves for the four QSAR approaches considered in this study and activity thresholds obtained when $pK_i(\text{II})$ (A) and $\Delta pK_i(\text{I-II})$ (B) are predicted.

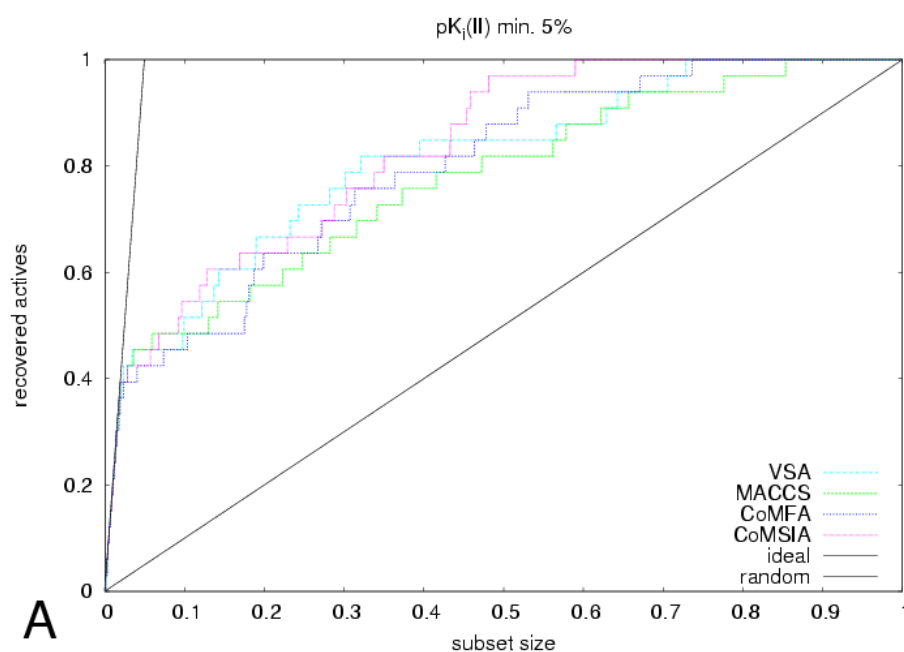
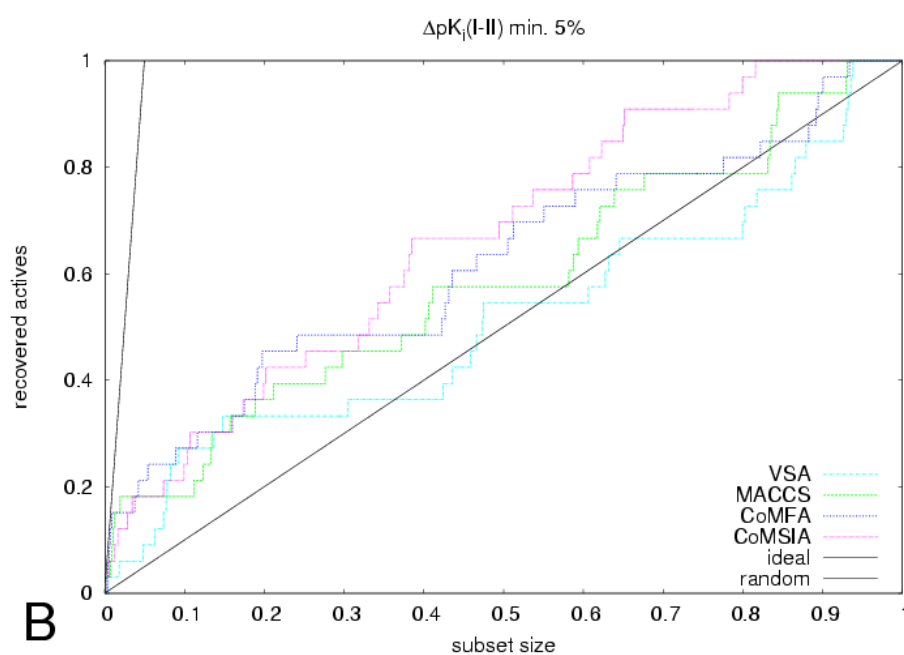
(a) $pK_i(\text{II})$ (b) $\Delta pK_i(\text{I} - \text{II})$

Figure 5.7: Enrichment plots for the retrieval of the 5% compounds with lowest $pK_i(\text{II})$ (A) and $\Delta pK_i(\text{I} - \text{II})$ (B), respectively, applying the four QSAR approaches considered in this study. The main diagonal corresponds to a random selection, the steep line to the left to an ideal retrieval.

models show a similarly satisfying albeit not perfect performance.

Figure 5.7(b) displays the enrichment curves for the retrieval of the 5% compounds with highest selectivity towards hCA II. In the very left hand side of the curves CoMFA, CoMSIA, and MACCS models discriminate satisfactorily for the first 1–2% of the database, whereas the VSA model does not depart from a random selection. For the remainder of the screening progress all models perform rather disappointingly, however, the 3D models still show some advantages. The VSA descriptors remain close to the performance of a random classifier.

5.5.7 Interpretation of the MACCS Models

Besides the use of a QSAR equation to predict novel compounds the coefficients of the equation indicate the relative importance of individual descriptors (usually multiplied by the standard deviation of the descriptor, denoted as $\text{stdev} \times \text{coeff}$). With respect to the evaluation of the 3D-QSAR models we refer to our recent paper [Hillebrecht et al., 2006] giving a detailed interpretation in terms of contour maps. Depending on the method applied, the resulting contours indicate either purely ligand-based, purely protein-based, or mixed protein-ligand-based information. In contrast to this study where an alignment has been produced exploiting the protein binding pocket as a reference, here we will use a similarity-based alignment strategy. Therefore, conclusions based on contour maps have to be interpreted with some caution, particularly referring to properties given by the protein environment. Nevertheless, since several crystal structures of complexes were used as reference some correlation between the properties of the binding pocket and the contour maps derived from the aligned ligands can be expected.

Each of the 166 MACCS descriptors captures the frequency of occurrence of a distinct molecular fragment in a molecule. Thus, if the product $\text{stdev} \times \text{coeff}$ adopts large absolute values the corresponding fragment takes strong influence on the biological activity. If the sign is positive its occurrence enhances biological activity (or selectivity, respectively). If it is negative, the presence of that fragment in a molecule is detrimental for its activity. Figure 5.8(a) displays a plot of 130 $\text{stdev} \times \text{coeff}$ values derived from the QSAR equation for $\text{pK}_i(\text{II})$, one for each MACCS key (36 of the 166 were discarded due to zero variance). The highest peak is experienced by MACCS_81, which corresponds to the fragment $\text{SA}(\text{A})\text{A}$ (where A is any atom except hydrogen). This pattern

is highly abundant in the dataset, but molecules exhibiting this fragment particularly frequently (two or three times, found in 25 molecules) achieve an extraordinarily high $pK_i(\text{II})$ ($\text{mean}(\text{subset}) = 8.63$, $\text{mean}(\text{all}) = 7.17$). The corresponding molecules possess in general more than one sulfonamide group mostly along with a thiophene ring. For example, the high-affinity thienothiopyranes (scaffold **A**) have one sulfonamide anchor, a second sulfonyl group in the thiopyrane ring and additionally a third time the pattern is matched by the thiophene ring. An example indicating affinity decrease is fragment MACCS_89, which encodes the number of OAAAO substructures. This pattern occurs particularly in sulfamates (scaffold **I**) which have significantly lower $pK_i(\text{II})$ values ($\text{mean}(\text{subset}) = 5.25$) compared to the entire training set. This is due to the replacement of the sulfonamide anchor by a sulfamate group, which is known to be a poorer zinc binder.

Figure 5.8(b) shows a plot of $\text{stdev} \times \text{coeff}$ derived from the QSAR equation for $\Delta pK_i(\text{I} - \text{II})$. The high value of the product $\text{stdev} \times \text{coeff}$ for MACCS_43 means that the substructure QHAQH (where Q is any atom except hydrogen and carbon) increases selectivity towards hCA I. The average for $\Delta pK_i(\text{I} - \text{II})$ is -1.29 (i.e. most compounds inhibit hCA II stronger than hCA I). Molecules with higher occurrence of this fragment compared to the remainder of the training set (46 molecules with more than one occurrence) hence possess a higher $\Delta pK_i(\text{I} - \text{II})$ ($\text{mean}(\text{subset}) = 0.03$). Compounds of this subset comprise structures with scaffold **H**, (thio)urea, and guanidinium moieties, respectively. An example for substructures increasing selectivity towards hCA II is MACCS_72 encoding OAAO. The average $\Delta pK_i(\text{I} - \text{II})$ for molecules containing this fragment more than once is -2.64. Hydroxysulfonamides (scaffold **F**) as well as ortho-dimethoxy substituted phenyl rings are captured by this pattern. A comparison of the $\text{stdev} \times \text{coeff}$ plots shows that many of the values are anticorrelated, i.e. the same MACCS key is associated with a positive sign for $pK_i(\text{II})$ and a negative sign for $\Delta pK_i(\text{I} - \text{II})$ and vice versa. This is a consequence of the fact that a factor increasing affinity towards hCA II often diminishes the difference $\Delta pK_i(\text{I} - \text{II})$.

These examples show that the MACCS-based models are not only useful for prediction of novel compounds, but that the coefficients of the derived equation can be used to identify substructures of importance to influence a molecular property such as activity. Of course, the conclusions drawn from this type of analysis are only of statistical nature. There is no basis to assume a causal relationship between occurrence of a distinct fragment and its effect on biological activity. It might well be, that, as a hypothetical example, a compound class such as thiophenes most frequently contains several sulfonyl

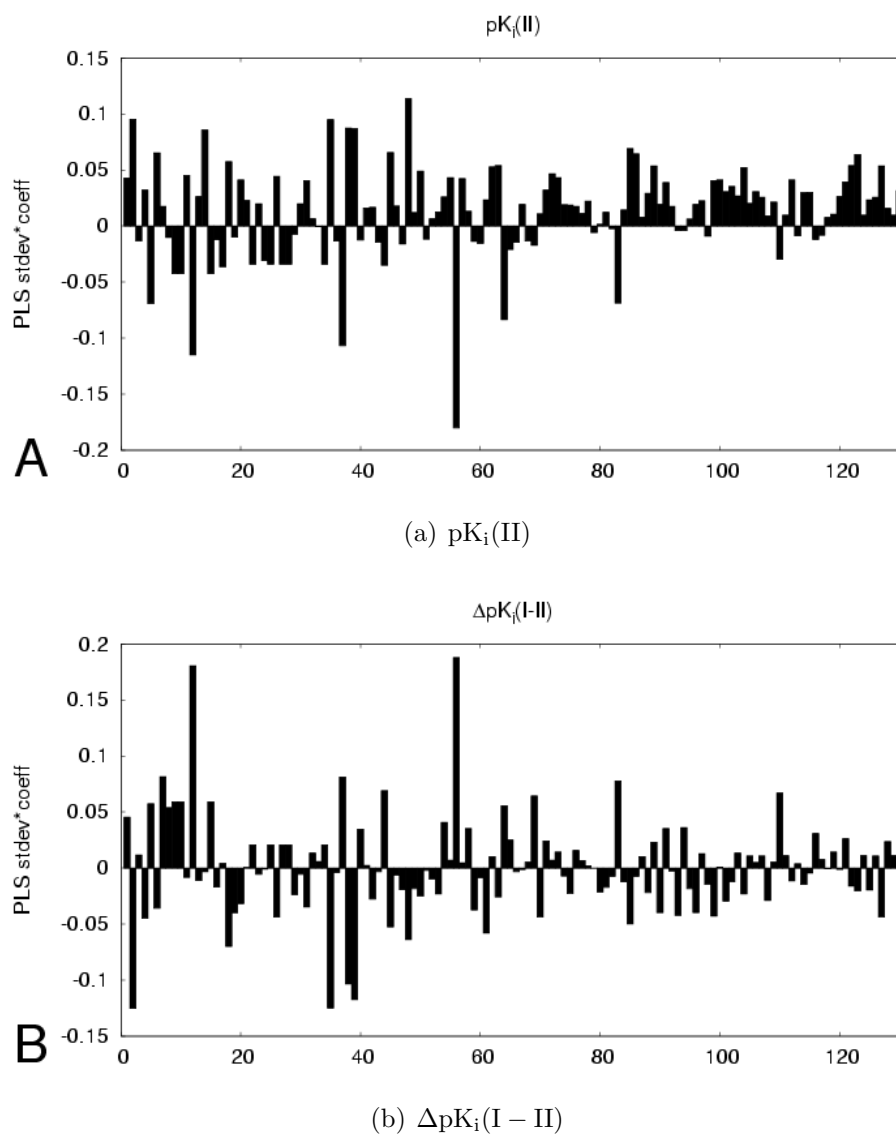


Figure 5.8: Bar plots of $\text{stdev} \times \text{coeff}$ derived from MACCS QSAR equations for prediction of $pK_i(\text{II})$ (A) and $\Delta pK_i(\text{I-II})$ (B), respectively.

groups and exhibits at the same time higher affinities than the average of the training set. However, it remains unresolved whether the thiophene ring, sulfonyl groups, or the combination of both are responsible for this observation. Since the MACCS keys do not take the functional characteristics of the encoded groups into account with respect to protein-ligand interactions, they can only indirectly evidence a possible correlation. Another limitation of this approach is that many of the encoded fragments are rather small and do not represent a meaningful “chemical unit”. The approach also does not consider any information about connectivity between single fragments. Thus, the method will fail if molecules are attempted to be predicted which possess a similar count of the same fragments as a subset of ligands from the training set but connected in a different way. Clearly, this shortcoming will not occur using a 3D-QSAR method as evaluation technique. Nevertheless, if narrow focused libraries with close similarity to the training set compounds are evaluated the MACCS key method will probably yield reasonable accuracy. An advantage of the fragmental description of the molecules is – despite ignorance of actual functional groups – the simple and straightforward translation into chemical structures. This is in contrast to many other property- or graph-based descriptors commonly applied in QSAR studies, e.g. the VSA descriptors used in our fourth approach to evaluate the data set. We therefore will refrain from a detailed interpretation. In principle the relevance of properties such as logP, MR etc. falling into distinct intervals could be assessed. However, the obtained information will remain rather indirect and general due to the global character of the descriptors. Particularly, it faces the problem of a non-trivial translation into chemical structures in accordance with the highlighted properties.

5.6 Conclusion

Within the present study we assessed the predictive power of QSAR approaches with respect to their applicability to screen databases. Since we were especially interested in 3D methods, a protocol had to be established which is reliable and robust enough to produce consistent spatial alignments of the molecules under consideration. Due to the large number of compounds encountered in real life screening scenarios the protocol – once set up – has to be applicable without further manual intervention. FlexS in combination with automated recognition of chemical compound class of ligands to be superimposed via the MAPREF methodology was chosen to successfully accomplish this task. We could demonstrate that CoMFA and CoMSIA models based on this

alignment perform comparably well with similar models based on manually derived alignment in the protein's binding pocket as a reference point along with subsequent force field relaxation. Since the superimposition comprises a rather elaborate and time consuming step we also tested the performance of alignment-free 1D- and 2D-QSAR models, particularly with respect to database screening. Therefore, fragment-based MACCS descriptors and property-based VSA descriptors were computed based on the 2D molecular information. The external predictivity was assessed based on a test set of 663 compounds with known activities. Of course, this number of test molecules does not touch the size of a real library but it should be sufficiently large for the intended benchmark test.

In summary, the 3D-QSAR models and the MACCS keys performed quite well with respect to affinity prediction ($\text{pK}_i(\text{II})$) whereas the VSA descriptors did not achieve to establish models with comparable predictive power. In terms of numerical affinity prediction the 3D-QSAR models significantly outperformed the 1D and 2D approaches. They tend to be more specific than the MACCS keys but at the cost of a lower sensitivity. The models are difficult to mutually rank against each other since relevance, predictive value, and applicability depend on the specific goal of the project, e.g. whether retrieval of only a few compounds with an enhanced activity is intended or whether as many actives as possible should be detected. Thus, even if the MACCS approach is easier to perform and computationally less demanding it does not make the 3D methods superfluous for screening purposes. Also with respect to the aspect of generality of the descriptors the molecular fields of a ligand approximate better the concept of molecular recognition and secondly they should not suffer from the fact of missing connectivity information as the MACCS keys do. Most likely, the 3D-QSAR models are much more robust in handling data sets composed of compounds with structurally rather diverse molecular skeletons. Nevertheless, we could show that the PLS coefficients of the derived MACCS model can be interpreted meaningfully and used to extract knowledge about the influence of individual fragments on the dependent variables.

With respect to a quantitative selectivity prediction, only the internal consistency was convincing. None of the techniques was able to make satisfying numerical forecasts on this large dataset. However, this can be attributed to the rather imbalanced composition of the test set and the generally higher error contained in a "composed variable". The results on categorical predictivity suggest, that the 3D models can still give crude estimates about selectivity.

The main purpose of this study was to systematically assess the performance of 3D-QSAR models with respect to database screening. Although the results are convincing one has to keep in mind that they will depend on the training and test set composition and cannot be transferred generally. Furthermore, additional investigations on other QSAR techniques (other statistical methods, 4D [Hopfinger et al., 1997], 5D [Vedani & Dobler, 2002], 6D [Vedani et al., 2005] approaches) need to be done to collect more experience on the scope and limitations of QSAR methods for database screening.

6 Bewertung virtueller Ligandbibliotheken durch QSAR-Modelle

6.1 Einleitung

In Kap. 5 wurde erörtert, inwieweit 1D-, 2D- und 3D-QSAR-Modelle zur Bewertung großer Moleküldatenbanken geeignet sind. Im Falle der 3D-QSAR-Modelle war zusätzlich die Entwicklung und Validierung eines robusten Protokolls zur Automatisierung des Alignments notwendig.

In diesem Kapitel wird eine mögliche praktische Anwendung anhand einer fokussierten virtuellen Ligandbibliothek gezeigt. Hierbei soll die Untersuchung im Sinne einer „*proof of concept*“-Studie“ durchgeführt werden. Die Erwartungen, die an ein solches QSAR-basiertes Datenbankscreening gestellt werden können, bestehen in der Auffindung strukturell modifizierter Verbindungen, deren Aktivitäten/Selektivitäten an den Extremen der Eigenschaften des Trainingsdatensatzes liegen. Anders ausgedrückt bedeutet dies, dass – eine Normalverteilung der vorherzusagenden Eigenschaften im Trainingsdatensatz vorausgesetzt – eine Anreicherung an den Extremen der Gausskurve möglich sein sollte. Es kann nicht erwartet werden, dass Moleküle mit völlig neuen Eigenschaften detektiert werden können, und es bleibt zu beachten, dass gerade an den niedrig populierte Extremen des Trainingsdatensatzes die Präzision der Vorhersagen abnimmt. Auch bezogen auf die chemische Struktur können nur Moleküle mit einer gewissen Ähnlichkeit zum Trainingsdatensatz bewertet werden. Als seriöses Ziel eines solchen QSAR-basierten Screenings kann also die Auffindung von strukturell abgeänderten Molekülen definiert werden, deren mittlere Aktivitäten/Selektivitäten signifikant über/unter dem entsprechenden Mittelwert des Trainingsdatensatzes liegen. Außerdem kann ein solcher Ansatz sehr wertvoll sein, da ggf. „vergessene Kombinationen“ von Gerüsten und Substituenten entdeckt werden können.

6.2 Allgemeine Kriterien für die Zusammensetzung der virtuellen Ligandbibliothek

Da als Bewertungsfunktion zum *Ranking* der zu entwerfenden Datenbank QSAR-Modelle verwendet werden sollen, muss darauf geachtet werden, dass der **Geltungsbereich** (engl. *applicability domain*) der angewendeten Modelle nicht überschritten wird. In erster Näherung kann dies realisiert werden, indem nur Verbindungen in die Datenbank aufgenommen werden, die eines der neun Grundgerüste (Abb. 4.1, **A–I**) enthalten. Ein weiteres Kriterium stellt die **einfache Synthetisierbarkeit** der Datenbankmoleküle dar. Um diese zu gewährleisten, wurden die Moleküle des Trainings- und Testdatensatzes aus Kap. 4.3.1 und 5.4.1 grob visuell auf ihr gemeinsames Bauprinzip analysiert. Im Wesentlichen lassen sie sich durch Reaktionen von Alkoholen oder Aminen mit Carbonyl- oder Sulfonylhalogeniden herstellen. Ferner wurde durch Beschränkung auf ein maximales Molekulargewicht der Einzelkomponenten sichergestellt, dass keine Moleküle mit hohem Molekulargewicht erhalten werden (ein Molekulargewicht von etwa 500 Da wurde angestrebt, aber nicht als feste Grenze verwendet). Weiterhin wurden *building blocks* nur in Datenbanken von Anbietern gesucht, die diese in größeren Mengen (Gramm-Maßstab) vertreiben, damit sie für eine spätere Synthese bezogen werden können.

6.3 Datenbankerzeugung

Die „*in silico*-Synthese“, also die Enummerierung der virtuellen Bibliothek, erfolgt mit FlexS^C, der kombinatorischen Erweiterung von FlexS. Zusammenbau von *building blocks* und Alignment finden unmittelbar hintereinander statt. Das automatisierte Alignmentprotokoll ist das gleiche wie in Kap. 5.4.1 beschrieben. Die entsprechenden *building blocks* werden zunächst durch eine einfache und schnelle 2D-Substruktursuche – codiert in Form eines SLN-strings (*SYBYL line notation*) – mit dem Programm UNITY aus einer Datenbank extrahiert. Zum Aufbau der Moleküle wurden folgende *building blocks* verwendet: 1508 Acylhalogenide mit einem Molekulargewicht ≤ 300 Da aus den Datenbanken ACD, ACROS, SIGMA und MAYBRIDGE_BUILDING_BLOCKS, wobei nur solche Moleküle extrahiert wurden, die genau eine Acylhalogenidgruppe besitzen; 996 Sulfonylchloride aus denselben Datenbanken mit einem Molekulargewicht ≤ 350 Da, die nur eine Sulfonylchloridgruppe aufweisen, und 2478 primäre aliphati-

sche Amine mit einem Molekulargewicht ≤ 300 Da, die genau eine primäre aliphatische Aminogruppe besitzen und aus der SIGMA-Datenbank stammen. Um das kombinatorische Modul von FlexS anwenden zu können, ist es notwendig, einen sog. *core* und eine oder mehrere R-Gruppen zu definieren. Damit das Programm das Verknüpfungsschema richtig realisiert, muss der Benutzer im *core*-Fragment ein oder mehrere Atome als sog. R-Atome (mit „R1, R2,...“ etc.) kennzeichnen. In den *building blocks* muss ein Atom als sog. X-Atom gekennzeichnet werden. Diese technischen Aufgaben wurden mit einem selbstgeschriebenen SPL-Skript bewerkstelligt, welches auf einem 2D-Substrukturmatching basiert.

In den folgenden Abschnitten sollen die entsprechenden Prozeduren für die einzelnen chemischen Klassen kurz erläutert werden.

6.3.1 para-Substituierte Phenylsulfonamide – Grundgerüst B

Zur Generierung para-substituierter Phenylsulfonamide wurden vier *core*-Strukturen verwendet: 4-(Aminosulfonyl)benzoylchlorid, 4-(Aminosulfonyl)benzolsulfonylchlorid, 4-(Amino)benzolsulfonsäure und 4-(Hydroxy)benzolsulfonsäure. Die Chloride werden mit primären aliphatischen Aminen zu entsprechenden Amiden verknüpft, das Amin und das Phenol werden mit Acyl- und Sulfonylhalogeniden zu Amiden bzw. Estern verbunden.

6.3.2 Thiadiazolsulfonamide – Grundgerüst A

Die Thiadiazolsulfonamide wurden *in silico* erzeugt, indem 5-Amino-3,4-thiadiazol-2-sulfonamid als *core* mit Acylhalogeniden oder Sulfonylchloriden zu entsprechenden Amiden verbunden wurden.

6.3.3 Benzothiazolsulfonamide – Grundgerüst C

Zur Erzeugung von Benzothiazolsulfonamiden wurden 6-(Amino)benzothiazol-2-sulfonamid und 6-(Hydroxy)benzothiazol-2-sulfonamid mit oder ohne einem Ethylenglykol-*Spacer* an der Hydroxy-Gruppe als *core* verwendet. Diese wurden mit Acylhalogeniden oder Sulfonylchloriden zu entsprechenden Estern und Amiden verknüpft.

6.3.4 Grundgerüste H und I

Die Verknüpfung von Sulfamid bzw. Sulfaminsäure mit Sulfonylchloriden ergibt die Verbindungen mit den Grundgerüsten **H** und **I**.

6.4 Bewertung der virtuellen Bibliothek mit QSAR-Modellen

Insgesamt wurden durch die oben genannten Prozeduren 23440 Moleküle erhalten und überlagert. Zur Bewertung der Moleküle wurden das CoMSIA- und das MACCS-Modell herangezogen, da die beiden Verfahren methodisch unterschiedlich sind und die besten Ergebnisse (s. Kap. 5) lieferten.

Als Ziel wurde gesetzt, solche Moleküle zu identifizieren, die eine möglichst hohe Affinität zur CA II besitzen, sich gleichzeitig aber als möglichst selektiv gegenüber CA I erweisen. Da es sich um eine Validierungsstudie handelt und nur relativ wenige (um 30) Verbindungen synthetisiert werden können, soll nur ein Grundgerüst, das am häufigsten im Trainingsdatensatz auftretende, betrachtet werden. Daher werden nur Phenylsulfonamide (Grundgerüst **B**) einbezogen. Um geeignete Grenzwerte für einen „hohen“ $pK_i(II)$ und einen „niedrigen“ $\Delta pK_i(I - II)$ festzulegen, wurde in der Gesamtheit der zur Verfügung stehenden Phenylsulfonamide ein Grenzwert ermittelt, der denjenigen 15% der Moleküle mit maximalem $pK_i(II)$ bzw. minimalem $\Delta pK_i(I - II)$ entspricht. Ferner wurde empirisch eine maximal zulässige Differenz zwischen den Vorhersagen beider Verfahren festgelegt. Dadurch soll eine Art „consensus-QSAR“ realisiert werden unter der Annahme, dass ein Ergebnis in Übereinstimmung zweier unterschiedlicher Methoden als zuverlässiger anzusehen ist.

Zusammenfassend wurden die Moleküle nach folgenden Kriterien ausgewählt:

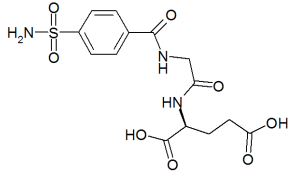
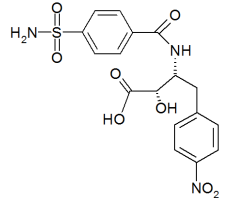
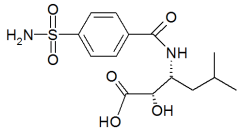
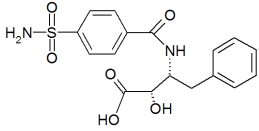
- Phenylsulfonamid-Grundgerüst (Gerüst **B**)
- $pK_{i,CoMSIA}(II) \geq 8.22$
- $\Delta pK_{i,CoMSIA}(I - II) \leq -2.07$
- $|pK_{i,CoMSIA} - pK_{i,MACCS}| \leq 0.7$ (gilt auch für Selektivitätsdifferenzen)

Insgesamt erfüllten 228 Moleküle diese Kriterien, daraus wurden 29 Kandidaten als Synthesevorschläge ausgewählt.

6.5 Ergebnisse

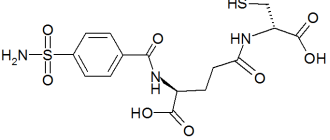
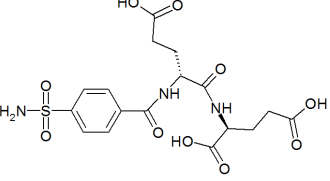
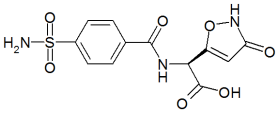
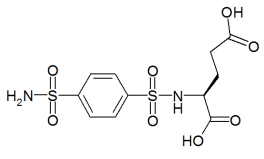
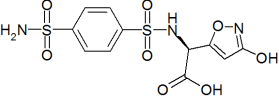
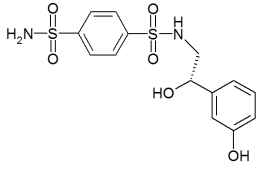
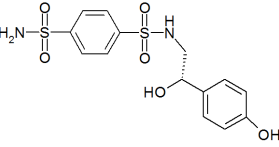
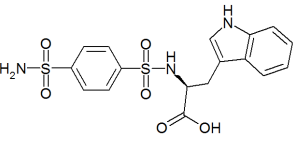
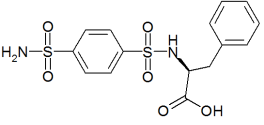
Die Strukturformeln der 29 ausgewählten Kandidaten sowie vorhergesagter $pK_i(\text{II})$ - bzw. $\Delta pK_i(\text{I} - \text{II})$ -Wert (für MACCS und CoMSIA) sind in Tab. 6.1 aufgelistet. Die Spalte T_c gibt den Tanimoto-Koeffizient des Moleküls an, der gegen das ähnlichste Molekül der gesamten CAI-Datenbank berechnet wurde (diese enthält neben dem Trainingsdatensatz zahlreiche weitere Moleküle, die in keiner der Studien für die vorliegende Arbeit verwendet wurden. Als Moleküldeskriptor wurden hierbei die UNITY 2D-Fingerprints und die *AtomPair*-Fingerprints zugrunde gelegt.

Tabelle 6.1: Strukturformeln der 29 Synthesekandidaten. Zu den Formeln sind die vorhergesagten Werte $pK_i(\text{II})$ und $\Delta pK_i(\text{I} - \text{II})$ für das MACCS- und das CoMSIA-Verfahren sowie der Tanimoto-Koeffizient T_c , der gegen das ähnlichste Molekül der gesamten CAI-Datenbank berechnet wurde, angegeben.

Nr.	Struktur	T_c	$pK_i(\text{II})$		$\Delta pK_i(\text{I-II})$	
			MACCS	CoMSIA	MACCS	CoMSIA
24		0.94	7.57	8.27	-1.22	-1.79
25		0.72	7.90	8.39	-1.92	-2.16
26		0.81	7.62	8.30	-1.74	-2.13
27		0.81	7.80	8.23	-1.69	-1.95

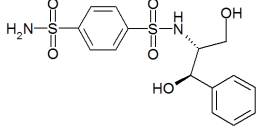
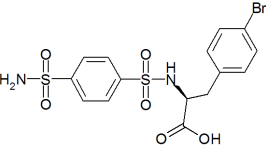
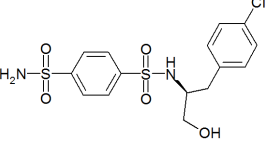
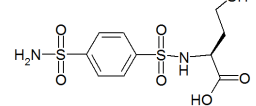
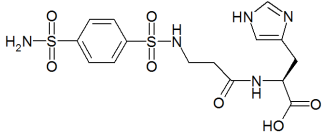
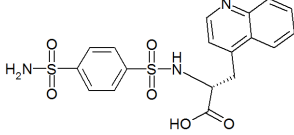
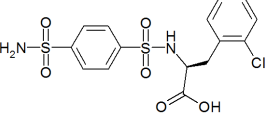
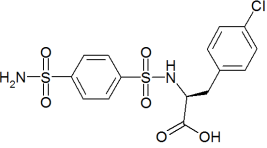
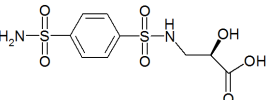
(Fortsetzung nächste Seite)

(Fortsetzung Tab. 6.1)

Nr.	Struktur	T_c	$pK_i(II)$		$\Delta pK_i(I-II)$	
			MACCS	CoMSIA	MACCS	CoMSIA
28		0.91	7.76	8.24	-1.26	-1.76
29		0.93	7.74	8.43	-1.22	-1.89
30		0.66	7.81	8.24	-2.01	-2.18
31		0.61	7.68	8.33	-1.69	-1.86
32		0.50	8.15	8.24	-1.81	-1.85
33		0.63	7.83	8.48	-1.81	-1.97
34		0.65	7.83	8.46	-1.81	-1.92
35		0.72	8.00	8.36	-1.34	-1.94
36		0.65	7.64	8.25	-1.67	-1.83

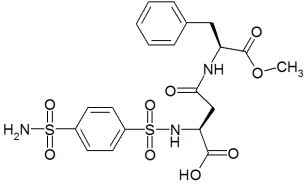
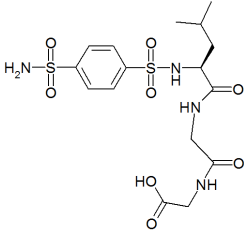
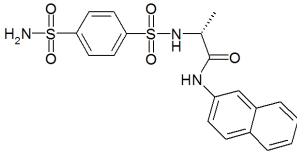
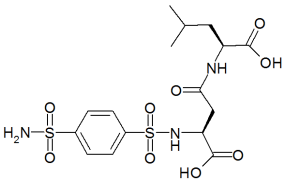
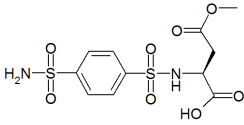
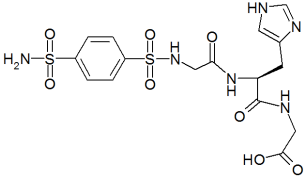
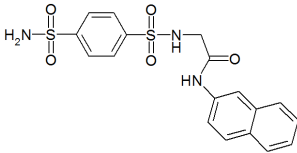
(Fortsetzung nächste Seite)

(Fortsetzung Tab. 6.1)

Nr.	Struktur	T_c	$pK_i(II)$		$\Delta pK_i(I-II)$	
			MACCS	CoMSIA	MACCS	CoMSIA
37		0.62	7.74	8.34	-1.62	-1.84
38		0.62	7.69	8.25	-1.83	-1.82
39		0.65	7.65	8.25	-1.66	-1.98
40		0.60	7.63	8.32	-1.83	-1.82
41		0.82	8.29	8.26	-1.43	-1.99
42		0.69	7.88	8.37	-1.48	-1.79
43		0.63	7.77	8.24	-1.68	-1.77
44		0.65	7.70	8.25	-1.59	-1.82
45		0.57	7.88	8.42	-2.05	-1.78

(Fortsetzung nächste Seite)

(Fortsetzung Tab. 6.1)

Nr.	Struktur	T_c	$pK_i(II)$		$\Delta pK_i(I-II)$	
			MACCS	CoMSIA	MACCS	CoMSIA
46		0.69	7.92	8.33	-1.52	-1.81
47		0.71	7.70	8.32	-1.50	-1.90
48		0.70	7.65	8.23	-1.31	-1.79
49		0.68	7.82	8.39	-1.61	-2.02
50		0.60	7.62	8.24	-1.61	-2.02
51		0.81	8.48	8.44	-1.28	-1.82
52		0.78	7.65	8.2	-1.43	-1.98

Die Verbindungen **24** bis **30** sind 4-(Aminosulfonyl)aminobenzoesäurederivate, **31** bis **52** sind von der 4-(Aminosulfonyl)sulfonsäure abgeleitet. Bei den angefügten Aminen handelt es sich meist um Aminosäuren, so dass die Moleküle in vielen Fällen eine

negativ ionisierbare Carboxylgruppe tragen. Ferner enthalten die verknüpften Reste in vielen Fällen aromatische Bausteine und Hydroxylgruppen.

Der Tanimoto-Koeffizient T_c zeigt, dass viele der ausgewählten Kandidaten signifikante strukturelle Unterschiede zu den bereits charakterisierten Inhibitoren aufweisen. Das Ähnlichkeitsmaß nimmt für die 29 Moleküle Werte zwischen 0.94 und 0.50 an; es besitzt einen Mittelwert von 0.7 bei einer Standardabweichung von 0.11.

6.6 Ausblick

Die Synthese und biologische Testung sollen prinzipiell durch die Gruppe von Prof. Dr. C.T. SUPURAN an der Universität Florenz durchgeführt werden. Allerdings lagen zum Zeitpunkt der Anfertigung dieser Arbeit weder synthetisierte Liganden noch biologische Testung vor. Sollte dies in der Zukunft geschehen, so könnte man durch einen t -Test überprüfen, ob der Mittelwert von $pK_i(\text{II})$ des Trainingsdatensatzes signifikant geringer ist als der der Kandidatenmoleküle. Ein entsprechender Test kann auch für $\Delta pK_i(\text{I} - \text{II})$ durchgeführt werden. Sollte die Signifikanz bewiesen werden können, wären damit die Erwartungen (vgl. Kap. 6.1), die man realistischerweise an solch ein QSAR-basiertes *data-mining* stellen kann, erfüllt.

7 Virtuelles Screening nach neuartigen Inhibitoren der Peptiddeformylasen (PDF)

7.1 Struktur, Klassifizierung und katalytischer Mechanismus der PDF

Die Peptiddeformylasen (PDF, EC 3.5.1.88) gehören zur Gruppe der Metalloproteinasen und kommen in Prokaryoten sowie in Plastiden und Mitochondrien von Eukaryoten vor. PDFs katalysieren die hydrolytische Abspaltung des Formylrestes von **Formylmethionin**, also derjenigen Aminosäure, mit der in Prokaryoten, Mitochondrien und Plastiden neu synthetisierte Peptide beginnen. Abgesehen von diesen Spezialfällen, beginnen die in anderen Kompartimenten neu synthetisierten Proteine mit einem nicht-formylierten Methionin. Bei etwa zwei Dritteln der reifen Proteine wird dieses jedoch durch **Methioninaminopeptidasen** (MAP, EC 3.4.11.18) entfernt [Giglione et al., 2004]. Diese Abspaltung kann MAP allerdings nur bei nicht-formylierten Peptiden katalysieren; Peptide mit Formylmethionin (fMet) müssen daher zuvor durch PDFs de-formyliert werden. Der gesamte zur Methioninabspaltung führende Reaktionsprozess wird als *N-terminal methionine excision* (NME) bezeichnet und ist in Gln. 7.1 und 7.2 dargestellt:



Bei fast allen PDFs ist das für die Katalyse erforderliche Metallion ein Fe^{2+} -Ion; sie sind daher äußerst oxidationsempfindlich [Groche et al., 1998; Rajagopalan & Pei, 1998]. Allerdings bleibt die katalytische Aktivität der PDFs nahezu vollständig erhalten, wenn das Eisen-(II)-Ion durch Ni^{2+} [Groche et al., 1998; Ragusa et al., 1999] oder Co^{2+} [Rajagopalan et al., 2000] ersetzt wird. Dieses Vorgehen wird daher üblicherweise in experimentellen Arbeiten (Enzymkinetikkmessungen, Kristallstrukturbestimmung) mit PDFs angewendet. Eine Substitution durch Zn^{2+} hingegen führt zum völligen Verlust der PDF-Aktivität [Giglione et al., 2000a], mit zwei bisher bekannten Ausnahmen:

einer PDF aus der höheren Pflanze *Arabidopsis thaliana* (AtPDF1A) und einer aus dem Bakterium *Leptospira interrogans* (LiPDF).

PDFs werden derzeit in drei Klassen gemäß ihrer Sequenzidentität eingeteilt: PDF1 (A und B), PDF2 und PDF3. PDF1B kommt in allen gramnegativen Bakterien, in den Plastiden von Pflanzen und bestimmten Parasiten, den *Apicomplexa*¹, vor. PDF2 kommt in allen grampositiven Bakterien vor und unterscheidet sich von den PDF1 durch zwei charakteristische Insertionen [Giglione et al., 2000a; Giglione & Meinnel, 2001; Guilloteau et al., 2002]. Aufgrund des Vorkommens der PDF in Bakterien und anderen Krankheitserregern und der inzwischen widerlegten Annahme, dass diese nicht im Menschen vorhanden seien, wurde die PDF als optimales Target für antibakterielle [Giglione et al., 2000a; Yuan et al., 2001; Pei, 2001] und antiparasitäre [Meinnel, 2000] Therapien angesehen. Dies relativierte sich, als man PDFs auch in den Mitochondrien von Pflanzen, Tieren und dem Mensch entdeckte [Giglione & Meinnel, 2001; Giglione et al., 2000b]. Diese PDF1A kommt in den Mitochondrien fast aller Eukaryoten vor, so auch in denen des Menschen. Dies ist ein wichtiger Aspekt bei der Entwicklung von PDF-Inhibitoren (PDFI) als Antibiotika: Der potente PDFI **Actinonin** zeigt nicht nur antibakterielle Wirkung, sondern hemmt auch die Proliferation menschlicher Zellen. Daher muss eventuellen zytotoxischen Nebenwirkungen von PDFI besondere Aufmerksamkeit geschenkt werden, sofern sie nicht als antineoplastische Wirkstoffe eingesetzt werden sollen. Andererseits kann das Vorkommen von PDFs in Pflanzen auch als Ausgangspunkt für die Entwicklung neuartiger Herbizide genutzt werden [Giglione et al., 2000b; Dirk et al., 2001; Howard et al., 2004]. Tab. 7.1 gibt einen Überblick über Einteilung und Merkmale der verschiedenen PDFs.

Da im Rahmen dieser Arbeit die PDF im Apicoplast von *P. falciparum* (PfPDF1B) von besonderem Interesse ist, sollen an ihrem Beispiel die strukturellen Eigenschaften erläutert werden [Robien et al., 2004]. Zur Kristallstrukturbestimmung wurde aus Stabilitätsgründen die Co²⁺-haltige Form des nativ Fe²⁺ enthaltenden Enzyms verwendet und die Struktur mit einer Auflösung von 2.2 Å gelöst. Die Sekundärstruktur wird von drei größeren α -Helices, drei antiparallelen β -Faltblättern und einer kleineren α -Helix in der Nähe des N-Terminus gebildet (Abb. 7.1).

¹ Apicomplexa (auch *Sporozoen* genannt) sind zu den Eukaryoten gehörende Parasiten, die infektiöse Sporozysten oder Oozysten bilden und einen charakteristischen Generationswechsel durchlaufen. Ihr Name rührt von ihrem besonderen Plastid-Organell, dem **Apicoplast**, her. Bekannte Vertreter sind z. B. der Malaria-Parasit *Plasmodium* oder *Toxoplasma gondii*, Erreger der Toxoplasmose.

Tabelle 7.1: Einteilung und Merkmale der Peptiddeformylasen. (Daten entnommen aus [Fieulaine et al., 2005])

Typ	Klasse	Vorkommen	Organismen mit verfügbarer 3D-Struktur	PDB-code
1	A	Mitochondrien	<i>Arabidopsis thaliana</i>	1ZX1
	B	alle gramnegative und einige grampositive Bakterien	<i>E. coli</i> , <i>Bacillus cereus</i>	1BS7, 1WS0
		Plastiden von Pflanzen und <i>Apicomplexa</i>	<i>P. falciparum</i>	1RL4
2		alle grampositive Bakterien	<i>B. stearothermophilus</i> ,	1LQY,
			<i>S. aureus</i> , <i>Streptococcus</i>	1LM4,
			<i>pneumoniae</i>	1LM6
3		Trypanosomatiden	— ^{a)}	— ^{a)}

^{a)}nicht bekannt

Bei den PDFs wird das Metallion (meist Fe^{2+}) in tetrahedraler Geometrie von zwei Histidinresten, einem Cystein und einem Wassermolekül koordiniert. Wie bei den CAs bewirkt die Bindung des Wassermoleküls an das Metallion eine Erniedrigung des pK_a -Wertes und erleichtert so die Deprotonierung und damit die Bildung eines nukleophilen Hydroxidions. In Abb. 7.2 ist ein Modell für den katalytischen Mechanismus der Deformylierung durch PDFs dargestellt, der mit experimentellen Befunden übereinstimmt [Becker et al., 1998].

Die Nummerierung der Aminosäuren bezieht sich in der folgenden Beschreibung auf die PDF aus *E. coli*. Im Ausgangszustand (Schritt 1) ist das deprotonierte Wassermolekül W1 (liegt also als Hydroxidion vor) mit dem Sauerstoff koordinativ an das Metallion gebunden und bildet mit seinem H-Atom eine Wasserstoffbrücke zu Glu133 aus. Ein zweites Wassermolekül W2 bildet mit dem einen H-Atom eine Wasserstoffbrücke zum Sauerstoff von W1, mit dem anderen eine zur Hauptketten-Carbonylgruppe von Gly45 aus. Das Sauerstoffatom von W2 bildet eine Wasserstoffbrücke zur NH-Gruppe des Leu91 und eine weitere zum Ne von Gln50 aus. Dies entspricht dem kristallographisch beobachtbaren Zustand von PDF-Apostrukturen.

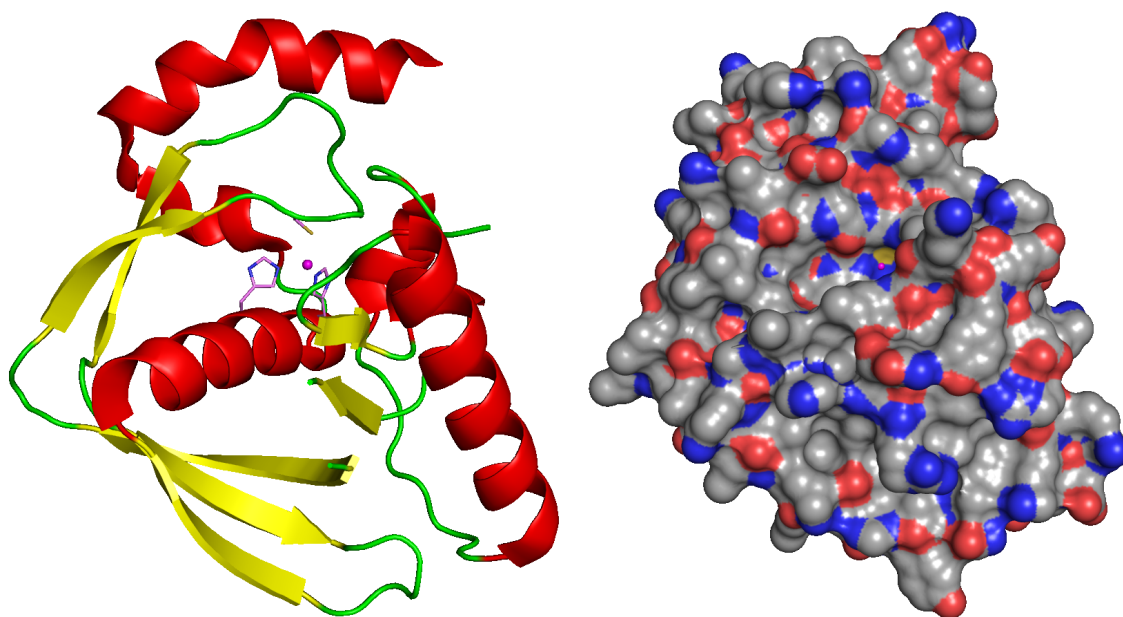


Abbildung 7.1: Struktur der Peptiddeformylase von *P. falciparum*. Links: Darstellung der Sekundärstrukturelemente. In der Mitte ist das Co^{2+} -Ion als Kugel in violett zu sehen, es wird von zwei Histidinresten und einem Cystein (ebenfalls in violett) koordiniert. Helikale Bereiche sind in rot, Faltblattstrukturen in gelb, Schleifenbereiche in grün dargestellt. Rechts: Darstellung der van-der-Waals-Oberfläche, eingefärbt nach den darunterliegenden Elementen (grau: C, rot: O, blau: N, gelb: S). Auch hier ist das Cobaltion als Kugel in violett zu sehen. (PDB-Code: 1RL4, in der Abb. ist der Ligand entfernt worden)

Im nächsten Schritt bildet sich der Enzym-Substrat-Komplex, wobei W2 durch ein Formylpeptid verdrängt wird. Der Carbonylsauerstoff der Formylgruppe bildet die entsprechenden Wasserstoffbrücken zu Leu91 und Gln50 aus; dadurch wird die Carbonylgruppe polarisiert und auf den nukleophilen Angriff durch das Hydroxidion vorbereitet. Durch diesen Angriff wird der Übergangszustand (Schritt 3) gebildet. Das Carbonylsauerstoffatom befindet sich im Zentrum eines Tetraeders, dessen Ecken durch das Metallion, den Carbonylkohlenstoff, das Hauptketten-N-Atom von Leu91 und das N_ϵ von Gln50 besetzt sind. In diesem Übergangszustand liegt das Metallion pentakoordiniert vor.

Das Proton von W1 wird mit Hilfe von Glu133 auf die Amidbindung des formylierten N-Terminus des Peptids übertragen, so dass der geladene Stickstoff zu ei-

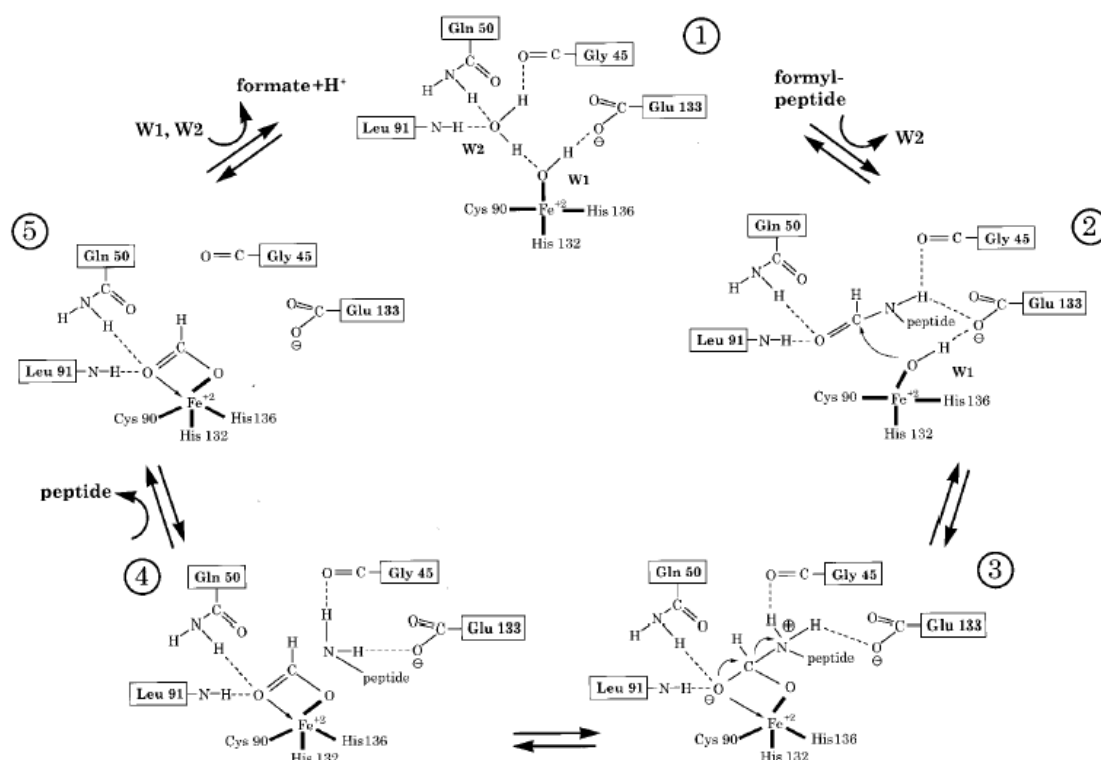


Abbildung 7.2: Modell des katalytischen Mechanismus der Peptiddeformylasen.

Die Nummerierung der Aminosäuren bezieht sich auf die PDF aus *E. coli*. Weitere Erklärung im Text. (Abb. entnommen aus [Becker et al., 1998])

ner guten Abgangsgruppe wird. Die folgende C-N-Bindungsspaltung führt zu dem ternären Enzym-Formiat-Peptid-Komplex (Schritt 4). Durch Freisetzung des Peptids entsteht der Enzym-Formiat-Komplex (Schritt 5). Dieser Zwischenzustand ist in Übereinstimmung mit biochemischen Befunden, die zeigen, dass PDFs Formylgruppen von einem auf ein anderes Peptid übertragen können. Der Reaktionszyklus schließt sich durch Aufnahme von zwei Wassermolekülen und Freisetzung von Formiat und einem Proton.

7.2 Bekannte Inhibitoren der PDF aus der Literatur

Im folgenden Abschnitt werden einige Informationen über bereits bekannte Inhibitoren der PDF zusammengestellt. Diese Information soll dann in die Aufstellung von Pharmakophorhypothesen einfließen. Beispielhaft werden auch einige Strukturformeln

und Affinitäten/Aktivitäten genannt, für vollständige Auflistungen sei auf die zitierten Übersichtsartikel verwiesen.

Alle PDF-Inhibitoren (im Folgenden PDFI abgekürzt) besitzen einen geeigneten Metallanker, der an das Fe^{2+} -Ion bindet. Eine Hydroxamatgruppe erweist sich hierbei als optimal, wie aus SAR-Studien homologer Verbindungen unter Variation des Metallankers hervorgeht [Boularot et al., 2004; Smith et al., 2002]. Eine Formylhydroxylamingruppe (also ein inverses Hydroxamat) besitzt eine verringerte Aktivität. Weitere mögliche Eisenankergruppen sind Hydrazide, Carbonsäurederivate, N-Hydroxyl(thio)harnstoffe, Thiole, Phosphatderivate, Oxime u.v.a.m. In Tab. 7.2, S. 150 sind beispielhaft die Strukturen und Affinitätsdaten von PDFI verschiedener chemischer Klassen dargestellt.

Prototyp der PDFI ist der Naturstoff **Actinonin (53)**, welcher von Actinomyceten produziert wird. Die peptidische Struktur des Inhibitors erklärt sich durch die Notwendigkeit, formylierte Polypeptide als natürliche Substrate der PDF zu imitieren. In der P_1' -Position besitzt es eine Pentylkette; andere Inhibitoren haben dort meist eine Butylkette (z. B. **54**, **55**) in Analogie zur viergliedrigen Seitenkette des Methionins, welches natürlicherweise die S_1' -Tasche besetzt. Diese Subtasche ist auch die am deutlichsten definierte und abgegrenzte Bindetasche von allen drei Subtaschen der PDF. Aus funktioneller Sicht ist dies auch sinnvoll, da das Enzym sämtliche mit Formylmethionin beginnenden Peptide prozessieren muss, dabei sind die auf das Formylmethionin folgenden Aminosäuren beliebig. Außer Pentyl- und Butylseitenketten findet man auch Cyclopentylethyl- (**56**) oder Phenylalkylseitenketten (**57**) als mögliche P_1' -Substituenten. Eine Besonderheit des Actinonin ist seine endständige Hydroxylfunktion, die als Donor mit der PDF eine zusätzliche Wasserstoffbrücke ausbilden kann. Die Kristallstruktur von Actinonin wurde im Komplex mit den PDFs verschiedener Organismen (*Pseudomonas aeruginosa*, *Escherichia coli*, *Bacillus stearothermophilus*, *Staphylococcus aureus*, *Leptospira interrogans*, *B. cereus*) gelöst. Der Bindungsmodus bleibt dabei nahezu unverändert.

Die Grundgerüste vieler weiterer PDF-Inhibitoren ähneln dem von Actinonin: Zahlreiche Derivate vom Typ des **VRC-4071 (55)** sind synthetisiert worden. Substitutionen in der α -Position des Hydroxamatankers brachten leichte Verbesserungen in der Affinität. Beim PDFI **BB-3497 (54)** ist der Metallanker ein inverses Hydroxamat, das Grundgerüst entspricht abermals dem Pseudopeptidtyp. Das **Thiorphan (57)** ist ein Beispiel für einen Inhibitor mit Thiolanker, welcher die relativ schwache Affinität zur

PDF erklärt. Das Molekül besitzt noch eine Peptidbindung und eine terminale Carboxylgruppe, so dass man es als „trunkiertes“ Peptid bezeichnen kann.

Eine weitere Abstraktion des Peptidgerüsts ist bei den Verbindungen vom Typ der β -Sulfinyl- oder β -Sulfonylhydroxamate realisiert worden. Ein Vertreter davon ist **Ro-66-0376 (58)**; hier übernimmt ein Sauerstoffatom der Sulfonyl- (oder Sulfoxid-) Gruppe die Akzeptorfunktion, die sonst von einer entsprechenden Carbonylgruppe ausgefüllt wird. Bei den **makrozyklischen Inhibitoren**, z. B. **59**, ist die peptidische Struktur noch deutlich zu erkennen, jedoch sind hierbei die Reste der S_1' - und der S_3' -Tasche kovalent miteinander verbunden. Eine weitere Leitstruktur für potente PDFI wird durch **bizyklische Inhibitoren** wie **60** repräsentiert. Bei ihnen übernimmt eine endozyklische Carbonyl- oder Sulfonylgruppe die Akzeptorfunktion des Carbonylsauerstoffs einer Peptidbindung des natürlichen Substrates. Der lipophile Phenylring ragt in die S_1' -Tasche.

Neben diesen wichtigeren Gerüsten sind einige andere Klassen mit meist niedrigerer PDF-Aktivität entdeckt worden, als Beispiel seien die thyroxinähnlichen Diphenylether (**61**) angeführt.

Bezüglich der Datenlage im Hinblick auf die Peptiddeformylase von *P. falciparum* ist anzumerken, dass derzeit nur für zwei Inhibitoren die IC_{50} -Werte publiziert wurden. Auffällig ist dabei, dass die Affinität für die PfPDF1B deutlich geringer (Faktor mehr als 1000) ist als die für die sehr nahe verwandten PDF1B gramnegativer Bakterien wie *E. coli*.

7.3 Pharmakophormodellierung und Datenbanksuchen mit CATALYST

Der Begriff des Pharmakophors wird 1998 durch die IUPAC wie folgt definiert:

„A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response.“

Die Art der Repräsentation solcher *features* variiert von Programm zu Programm. In der vorliegenden Arbeit wurde das Softwarepaket CATALYST [CATALYST 4.11] zur

Generierung der Pharmakophormodelle und für die anschließenden Datenbanksuchen verwendet. Daher fasst dieser Abschnitt kurz zusammen, wie Pharmakophormodelle in CATALYST kodiert und zur Datenbanksuche verwendet werden.

Im Gegensatz zu Programmen wie UNITY [Martin, 1992] werden die Konformationen der zu durchsuchenden Moleküle nicht *während* der Datenbanksuche, sondern *vorher* erzeugt. Dies hat den Vorteil, dass die Suchen mit vorerzeugten Konformationen erheblich schneller sind, allerdings ist mehr Festplattenspeicherplatz erforderlich. Der Konformationsgenerator von CATALYST basiert auf dem CHARMM-Kraftfeld [Brooks et al., 1983]. Die Konformationen werden im Vakuum erzeugt, wobei ein sog. *Poling*-Algorithmus für eine möglichst breite Abdeckung des Konformationsraumes sorgt. Durch vergleichende Studien wurde gezeigt, dass kleine Moleküle bei der Bindung an ein Protein häufig Konformationen annehmen, die *nicht* einem globalen oder lokalen Minimum in Lösung oder im Vakuum entsprechen [Perola & Charifson, 2004]. Um also in dem erzeugten Konformationsensemble auch die bioaktive Konformation eines Liganden vorzufinden, ist es notwendig, auch solche Konformere zu erzeugen, die deutlich über dem vom Kraftfeld errechneten globalen Minimum liegen. Bei dem hier vorgestellten Screening wurden die Konformationen so erzeugt, dass sie bis zu 20 kcal/mol über dem globalen Minimum liegen. Pro Molekül wurde die maximal erzeugte Anzahl von Konformationen auf 250 begrenzt (es wurde die Einstellung „FAST“ zur Datenbankerzeugung verwendet, Details siehe Handbuch zu [CATALYST 4.11]).

Zur eigentlichen Notation eines Pharmakophores besitzt CATALYST eine Sammlung vorgegebener Funktionen und Fragmente, z.B. für hydrophobe Molekülteile (auch unterteilbar nach aliphatischen/aromatischen), Wasserstoffbrückendonoren und -akzeptoren oder positiv/negativ ionisierbare/geladene Gruppen. Diese sind in einem *feature dictionary* abgelegt und decken gängige Wechselwirkungsfunktionen bei Protein-Ligand-Interaktionen ab. Diese Funktionen werden entweder punktförmig beschrieben, umgeben von einer kugelförmigen Toleranzsphäre (z.B. hydrophobe Funktionen), oder als Vektorobjekte, um bei Richtungsabhängigkeiten wie Wasserstoffbrückenbindungen die Richtungsabhängigkeit zu berücksichtigen. Weiterhin ist es möglich, benutzerdefinierte Funktionen einzuführen; dies war bei dem hier durchgeführten Screening zur Definition eines Metallankers, für den verschiedene chemische Fragmente in Frage kommen, notwendig.

Bei der eigentlichen Datenbanksuche kann zwischen zwei Algorithmen gewählt werden: Bei der Einstellung „FAST“ wird versucht, die Moleküle durch Translation und Rota-

tion in den Pharmakophor einzupassen, an den Torsionen werden keine Änderungen vorgenommen. Dieser Algorithmus ist relativ schnell, allerdings kann es v.a. bei Molekülen mit vielen rotierbaren Bindungen vorkommen, dass diese aufgrund der begrenzten Abdeckung des Konformationsraumes fälschlicherweise nicht den Pharmakophor passieren. Wird die Einstellung „BEST“ gewählt, versucht das Programm zusätzlich, durch leichte Änderungen der Torsionwinkel das Einpassen in den Pharmakophor zu ermöglichen. Dabei kann vom Benutzer ein Energiebetrag vorgegeben werden, den keine Konformation beim Variieren der Torsionswinkel überschreiten darf. In der vorliegenden Datenbanksuche wurde der BEST-Algorithmus mit einem Standardwert von 9.5 kcal/mol als maximalem Energieanstieg verwendet.

Das Ergebnis einer CATALYST-Datenbanksuche ist einerseits eine Information, welche Moleküle den Pharmakophor erfüllen und andererseits ein sog. „FIT“-Wert, der angibt, wie weit die *features* eines Moleküls von den Idealpositionen des vorgegebenen Pharmakophors abweichen. Höhere FIT-Werte bedeuten dabei eine geringere zugelassene Abweichung vom Pharmakophor. Als weiteres Kriterium kann die Energie des am besten in den Pharmakophor passenden Konformers ausgegeben werden; damit können allzu stark verzerrte Geometrien identifiziert werden.

7.4 Generierung von Pharmakophorhypothesen für PDF-Inhibitoren

Um Pharmakophormodelle zu erzeugen, müssen zunächst essentielle und optionale *features* identifiziert werden. Dies erfolgte hierbei durch Überlagerung von PDF-Kristallstrukturen, die einen Liganden enthalten. Dabei wurden nur diejenigen Strukturen berücksichtigt, die Liganden enthalten, welche direkt an das Metallion koordinieren. Tabelle 7.3, S. 151 gibt eine Auflistung dieser Kristallstrukturen.

Abbildung 7.3 zeigt die Überlagerung der Liganden auf die Proteinstruktur der *P. falciparum*-PDF (PDB-Code: 1RL4).

Folgende pharmakophore Elemente lassen sich erkennen:

- Alle Inhibitoren enthalten einen geeigneten **Fe²⁺-Anker** (Hydroxamat, inverses Hydroxamat, Phosphat).

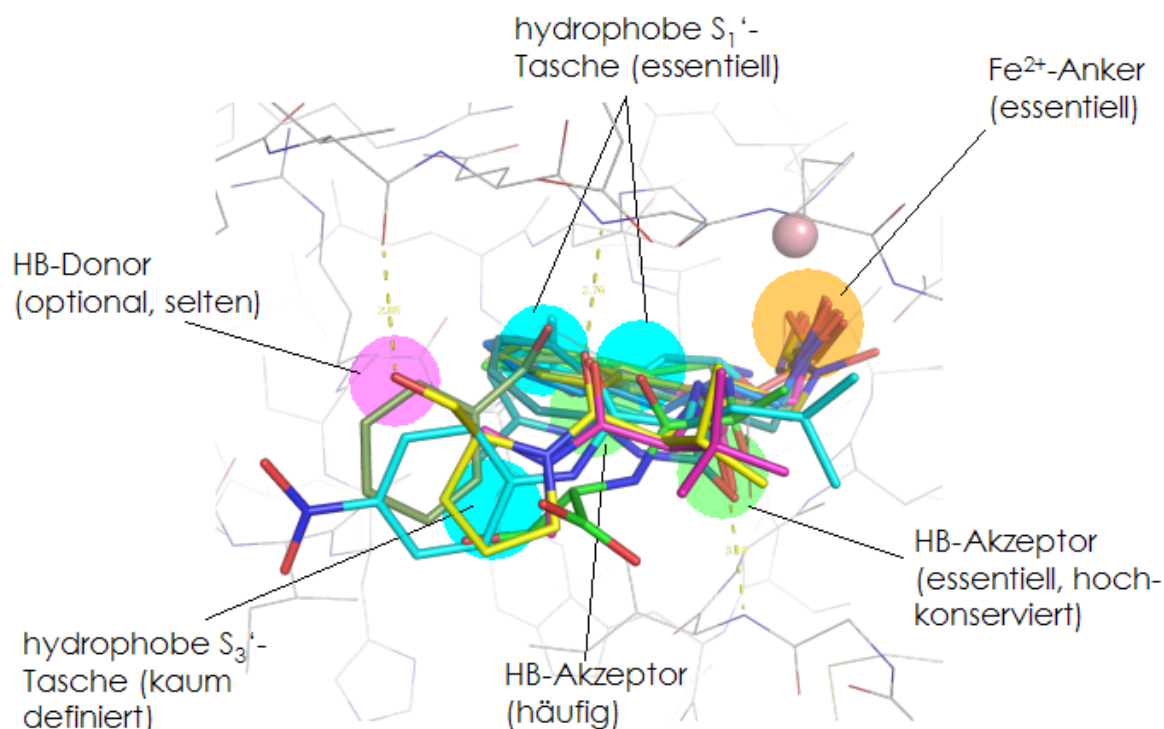


Abbildung 7.3: **Ableitung eines Pharmakophormodells für PDF-Inhibitoren.** In dünner, grauer Linienrepräsentation ist die PDF-Struktur aus *P. falciparum* (PDB-Code: 1RL4) dargestellt, das Eisen-(II)-Ion ist als rötliche Kugel zu sehen. Die Inhibitoren aus anderen PDF-Strukturen wurden überlagert und sind in farbiger Stabrepräsentation dargestellt. Zur besseren Übersicht ist nur eine der in Tab. 7.3 aufgeführten Actinoninstrukturen abgebildet.

- Ein **hydrophober Rest**, der die **S₁'-Tasche** ausfüllt (Butyl, Alkylaryl, Met-Seitenkette).
- Die meisten Liganden besitzen einen **Wasserstoffbrückenakzeptor** (Carbonylgruppe), der eine Wasserstoffbrücke zu Ile105 (Nummerierung gemäß *P. falciparum*-PDF) ausbildet.
- Viele der potenteren Inhibitoren besitzen einen **zweiten Wasserstoffbrückenakzeptor**, der eine Wasserstoffbrücke zu Gly154 ausbildet.
- Nur bei den größeren Inhibitoren ist ein **zweiter hydrophober Rest** vorhanden, der die wenig definierte **S₃'-Tasche** adressiert. Dieser Bereich ist relativ stark solvensexponiert, so dass ein solches *feature* eher geringere Bedeutung besitzt.
- Eine zusätzliche **Donorfunktion** ist einzig beim Actinonin zu finden. Die Hydro-

xylgruppe bildet eine Wasserstoffbrücke zur Carbonylgruppe der Hauptkette von Ile152 aus. Dieses *feature* macht einen Pharmakophor sehr selektiv, allerdings ist zu erwarten, dass nur sehr wenige Moleküle dieses *feature* besitzen.

Um die genaue Lokalisation der *features* festzulegen, wurde Actinonin in seiner bioaktiven Konformation aus der Kristallstruktur im Komplex mit der PDF von *E. coli* (PDB-Code: 1G2A, Auflösung: 1.8 Å) extrahiert. Auf den funktionellen Gruppen dieser Konformation wurden die entsprechenden Pharmakophorelemente platziert. Dadurch ergibt sich ein Basispharmakophor, der in Abb. 7.4 dargestellt ist: Auf der linken Seite ist eine zweidimensionale Repräsentation von Actinonin und den entsprechenden Wechselwirkungen mit den Aminosäuren der *P. falciparum*-PDF gezeigt, auf der rechten Seite die entsprechende 3D-Illustration.

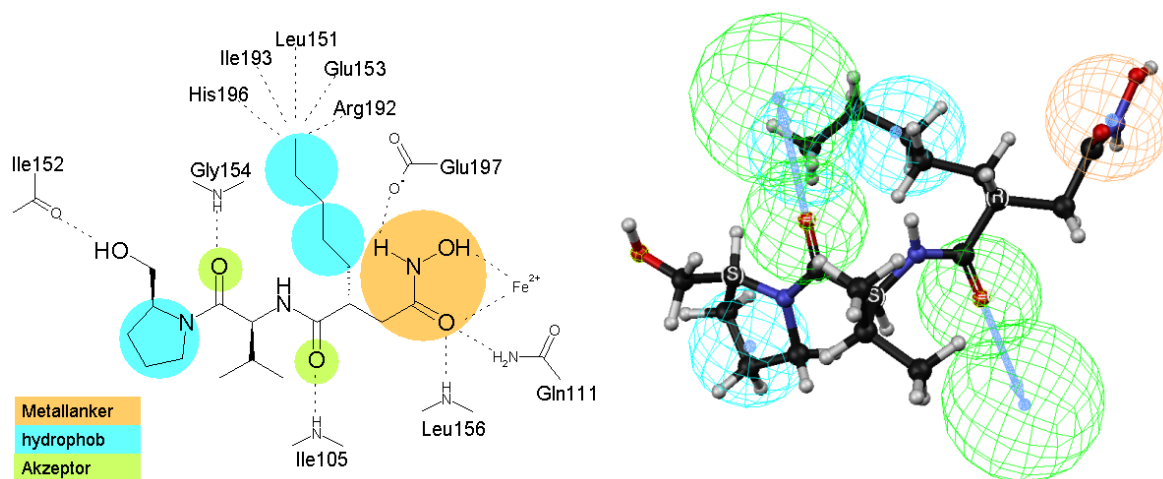


Abbildung 7.4: Ableitung eines Basispharmakophors von der bioaktiven Konformation des Actinonin (extrahiert aus PDB-Code: 1G2A). Der linke Teil zeigt schematisch in zweidimensionaler Darstellung das Wechselwirkungsmuster des Actinonin, im rechten Teil ist die entsprechende dreidimensionale Darstellung zu sehen.

Für alle *features* wurde eine Toleranz von 1.5 Å auf Ligandseite gewählt. Bei gerichteten *features* wurde für den auf der Proteinseite lokalisierten Projektionspunkt eine Toleranz von 2.0 Å definiert. Für den in orange dargestellten Metallanker wurde ein sog. *customized feature* generiert, das die in der Referenz [Boularot et al., 2004] als geeignete Fe^{2+} -Anker aufgeführten terminalen Gruppen enthält. Diese Fragmente sind in Abb. 7.5 aufgelistet und wurden zur Erzeugung des Metallanker-Features mit einer logischen ODER-Verknüpfung kombiniert, d.h. zur Erfüllung dieses *features* muss ein

Molekül eines dieser Fragmente enthalten, und das Fragment muss innerhalb der Toleranzsphäre lokalisiert sein. Zugunsten geringerer Laufzeiten während der Datenbank-suche wurde auf Einbeziehung von Direktionalitätsaspekten, die bei der Chelatisierung eines Metallions zweifelsohne eine Rolle spielen, verzichtet.

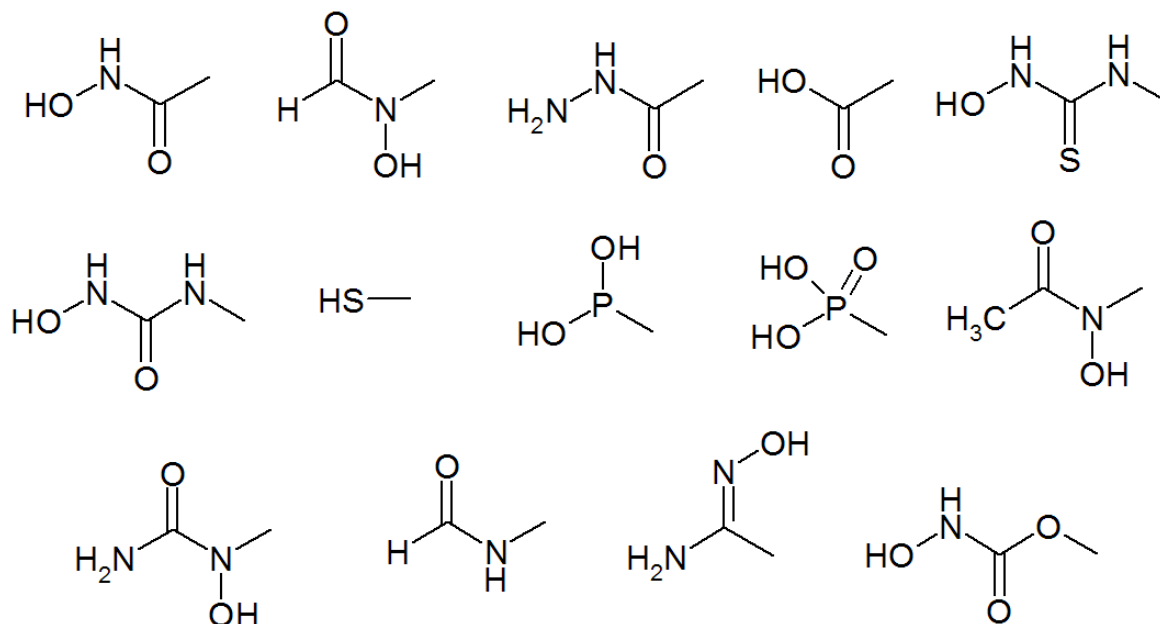


Abbildung 7.5: Liste der Fragmente, die zur Definition des Metallanker-Features verwendet wurden.

7.5 Selektivitätsmodellierung gegenüber humaner PDF

Lange wurde davon ausgegangen, dass im menschlichen Organismus keine PDFs exprimiert werden; neuere Untersuchungen zeigten jedoch, dass in den Mitochondrien eine humane HsPDF1A lokalisiert ist. Actinonin inhibiert auch diese PDF und wirkt dadurch zytotoxisch auf humane Zelllinien (s. Kap. 7.1). Deshalb ist es wichtig, entsprechende Selektivitätsbetrachtungen bei der Suche nach PDF-Inhibitoren zu berücksichtigen.

Eine Kristallstruktur der humanen HsPDF1A ist noch nicht aufgeklärt worden, daher wurde ein Homologiemodell aus dem SWISS-MODEL Repository [Kopp & Schwede, 2004] verwendet. Als Templat diente die Kristallstruktur der AtPDF1A aus den Mitochondrien der Pflanze *Arabidopsis thaliana* (PDB-code: 1ZXZ), die eine Sequenzidentität von 40% zur Aminosäuresequenz der humanen mitochondrialen PDF1A (Primary

accession number: Q9HBH1) aufweist. Abb. 7.6 zeigt eine Überlagerung des Homologie-modells der HsPDF1A (Antitarget) mit der Kristallstruktur der parasitären PfPDF1B (PDB-code: 1RL4, Target); in gelb ist der gebundene Inhibitor gezeigt, in grau die van-der-Waals-Oberfläche der PfPDF1B und in orange ausgewählte Aminosäurereste sowie Teile der Sekundärstruktur der HsPDF1A. Im wesentlichen bestehen zwei Unterschiede zwischen der HsPDF1A und der PfPDF1B (in Analogie zur AtPDF1A und EcPDF1B, wie in Ref. [Fieulaine et al., 2005] beschrieben): Die S_1' -Tasche ist im Modell der HsPDF1A durch die Aminosäuren Phe107 (Nummerierung gemäß AtPDF1A-Templat) und Trp146 klarer definiert und verengt (entsprechende Aminosäuren in PfPDF1B: Leu und Ile). Ferner ist die S_3' -Tasche durch eine charakteristische Insertion, den sog. CD-Loop, klarer begrenzt als bei der PfPDF1B.

Für eine Berücksichtigung der Selektivität bei der Suche mit geeigneten Pharmakophoren ist am ehesten das Trp146 heranzuziehen, da es am nächsten zur Bindetasche lokalisiert ist. Um den erhöhten Raumanspruch dieses Restes zur Gewinnung von Selektivität in eine Pharmakophorhypothese einfließen zu lassen, bietet es sich an, ein *feature* zu definieren, das durch ein beliebiges Ligandatom erfüllt wird (also ein „ANY-*feature*“ oder ein „included volume“). Zur Platzierung eines solchen *features* dient dann der Indolring des Trp146. Liganden, die ein solches Pharmakophor erfüllen, weisen einen sterischen Konflikt mit dieser Aminosäure auf, so dass die Affinität bezüglich des Antitargets HsPDF1A deutlich verringert sein sollte. Es muss allerdings darauf hingewiesen werden, dass solche Liganden aufgrund des Abstandes des Trp146 zum Metallanker eine gewisse Mindestgröße besitzen müssen.

7.6 Pharmakophorvalidierung und Durchführung der Datenbanksuchen

Um verschiedene Pharmakophorhypothesen zu testen, wurden 33 aktive PDF-Inhibitoren aus der Literatur extrahiert; dabei wurden möglichst strukturell diverse Verbindungen berücksichtigt. Jeder Pharmakophor wurde mit Hilfe dieser Moleküle auf seine **Sensitivität** getestet, d.h. auf seine Fähigkeit, aktive Inhibitoren zu erkennen bzw. als Filter passieren zu lassen. Parallel ist zu testen, ob der Pharmakophor ausreichende **Spezifität** besitzt, also möglichst keine als inaktiv bekannten Verbindungen passieren zu lassen. Die Zusammenstellung eines Datensatzes von verifizierten inaktiven Verbin-

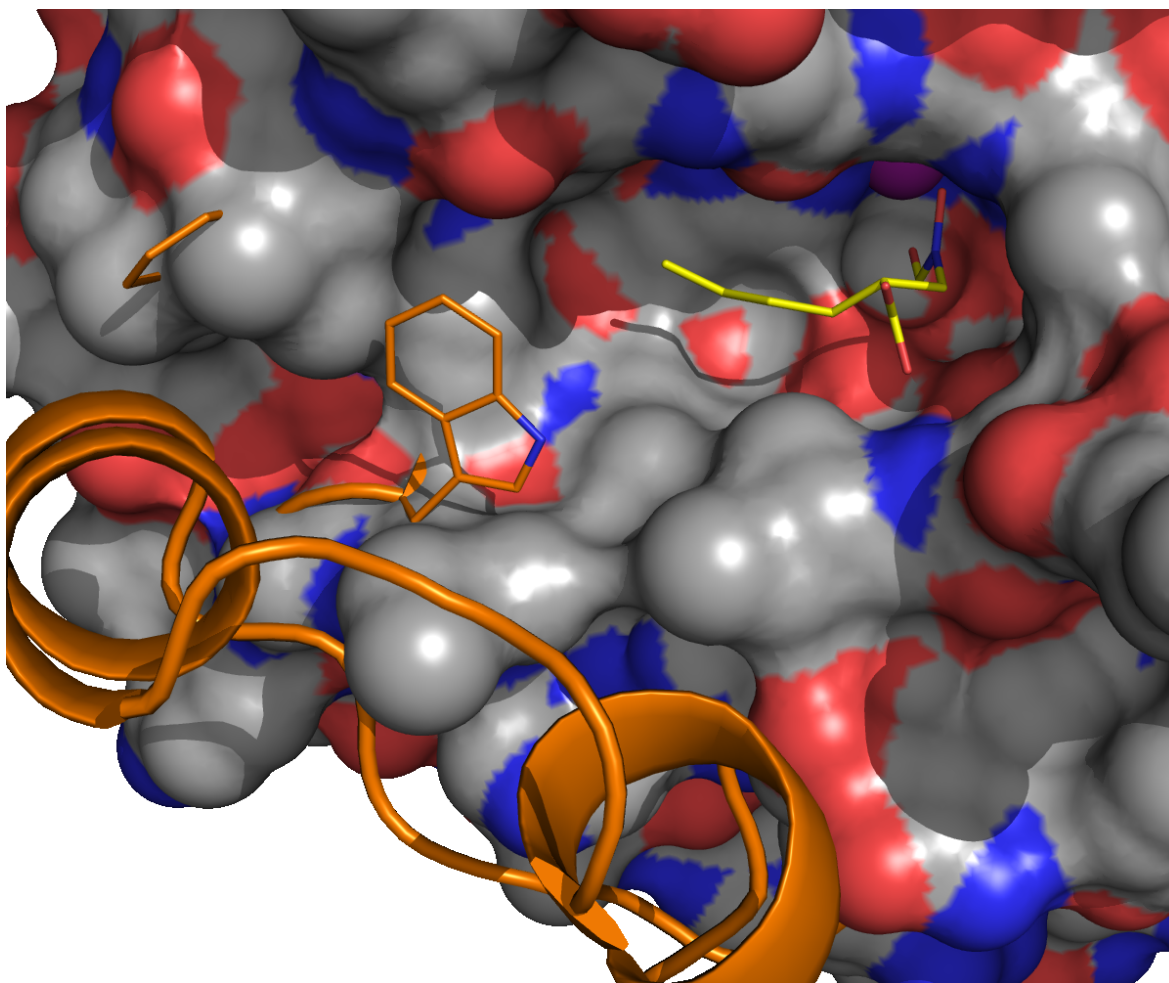


Abbildung 7.6: Überlagerung der Kristallstruktur von PfPDF1B mit einem Homologiemodell der HsPDF1A zur Identifizierung selektivitätsbestimmender Faktoren. PfPDF1B (PDB-code: 1RL4) ist mit der van-der-Waals-Oberfläche in grau dargestellt, der gebundene Inhibitor in gelb. Strukturen in orange gehören zur HsPDF1A; die beiden sterisch anspruchsvollen Reste Phe107 und Trp146 sind in der Nähe zur S_1' -Tasche zu sehen. Der helikale Bereich in Sekundärstrukturdarstellung ist eine charakteristische Insertion der mitochondrialen PDF1As. Er begrenzt die S_3' -Tasche.

dungen ist ein generelles Problem bei sog. *Enrichment*-Studien; theoretisch müssten alle als inaktiv klassifizierten Moleküle in einem entsprechenden Assay auf ihre Inaktivität experimentell überprüft werden. Aufgrund der hohen Anzahl (1000 bis 10000) von Verbindungen, die man für derartige Studien benötigt, wird in der Praxis häufig auf zufällig ausgewählte Verbindungen, welche die „Lipinski’s Rule of Five“ [Lipinski et al., 2001] erfüllen, zurückgegriffen. An dieser oft praktizierten Methode ist einerseits

zu bemängeln, dass die „Rule of Five“ sich nur auf die perorale Bioverfügbarkeit von Molekülen beziehen, jedoch ohne Bedeutung für die Wahrscheinlichkeit einer Wirkung an Rezeptoren sind. Zudem haben entsprechende Screening-Experimente gezeigt [Verdonk et al., 2004], dass die inaktiven Verbindungen ähnliche Verteilungen bezüglich eindimensionaler physikochemischer Eigenschaften aufweisen sollten wie die aktiven, um nicht unrealistisch hohe Anreicherungsraten (also eine zu optimistische Bewertung des Screeningprotokolls) zu erhalten. In der vorliegenden Datenbanksuche ist ein solches Vorgehen aber nicht angewendet worden; ein Metallanker ist notwendige Voraussetzung, damit die PDF inhibiert werden kann. Also wäre es sinnlos, von vorneherein inaktive Moleküle ohne Metallanker zu verwenden, denn dabei würde man nur die Leistung des Pharmakophors, Metallanker zu erkennen, messen. Würde man aber zufällig zusammengestellte Verbindungen mit Metallanker als inaktiv deklarieren, so wäre hierdurch die Annahme, dass diese Verbindungen mit hoher Wahrscheinlichkeit auch inaktiv sind, nicht zwingend tragbar, da alleine die Präsenz eines Metallankers schon die Wahrscheinlichkeit einer Inhibition signifikant erhöht.

Bei der hier durchgeführten Datenbanksuche wurde aus diesem Grund neben einer ausreichenden Sensitivität überprüft, inwieweit der getestete Pharmakophor in der Lage ist, die zu durchsuchende Datenbank kommerzieller Verbindungen befriedigend zu reduzieren, ohne hierbei Annahmen über die Aktivität der den Filter passierenden Moleküle zu antizipieren. Als zu durchsuchende Datenbank wurde die ZINC-Datenbank [Irwin & Shoichet, 2005] verwendet; es handelt sich um eine frei verfügbare Datenbank, die eine Zusammenstellung von Substanzen verschiedenster Anbieter enthält. Die Daten sind in aufbereiteter Form zu vordefinierten Untermengen nach bestimmten physikochemischen Eigenschaften zusammengefasst. Im Falle der vorliegenden Arbeit wurde das *druglike-subset* verwendet, das 2066906 Moleküle enthält (Stand: 2.5.2006). Hierbei sind die Moleküle nach folgenden Kriterien vorgefiltert: logP-Wert kleiner als 5, Molekulargewicht zwischen 150 und 500, Anzahl rotierbarer Bindungen kleiner als 8, polare Oberfläche kleiner als 150 und Zahl der Wasserstoffbrückenakzeptoren kleiner als 10. Diese Verbindungen wurden als SD-File heruntergeladen, mit dem Konformationsgenerator CORINA [Gasteiger et al., 1990] wurden die Atomtypen überprüft, 3D-Koordinaten erzeugt und ionisierbare Gruppen neutralisiert. Die so erhaltenen MOL2-files wurden in eine UNITY-Datenbank importiert. Eine 2D-Suche nach den in Abb. 7.5 aufgeführten potentiellen Metallankern reduzierte die Moleküle auf eine Zahl von 81498 (entsprechend 3.94% der Ausgangsdatensammlung), davon enthalten 76275 eine Carboxylgruppe, nur 5223 besitzen einen anderen Metallanker. Außerdem wurde der *World*

Drug Index (WDI) durchsucht; es handelt sich um eine Sammlung aus klinisch verwendeten Arzneistoffen sowie anderen Substanzen mit annotierter biologischer Aktivität (präklinische/experimentelle Substanzen, Pflanzenschutzmittel etc.). Er umfasst 67050 Verbindungen; auf eine Vorfilterung nach Metallankern wurde hierbei verzichtet. Zur Durchführung der eigentlichen Suchen wurden die Moleküle in einer CATALYST-Datenbank unter Anwendung der in Kap. 7.3 beschriebenen Parameter abgelegt.

Ausgehend von dem in Kap. 7.4 beschriebenen Basispharmakophor (Abb. 7.4, S. 129) wurden verschiedene Modifikationen getestet. Die Ergebnisse der Datenbanksuchen sind in Tab. 7.4, S. 152 aufgeführt; dabei sind jeweils die absoluten und relativen Zahlen der Moleküle angegeben, die die jeweilige Pharmakophorhypothese erfüllen. Diese Angaben sind für die 33 aktiven Inhibitoren aus der Literatur und für die Moleküle mit Metallanker aus der ZINC-Datenbank aufgelistet, wobei dort zwischen Molekülen mit (ZINC-COOH) und ohne (ZINC-noCOOH) Carboxylgruppe unterschieden wird. Ferner sind die Ergebnisse für das Screening des WDI aufgeführt.

Abb. 7.7 zeigt die fünf getesteten Hypothesen; jeder Hypothese wurde ein Buchstabe zugewiesen, mit dem der jeweilige Pharmakophor im folgenden Text und in Tabelle 7.4, S. 152 gekennzeichnet wird. Modell **A** ist der bereits beschriebene Basispharmakophor; als Startpunkt für weitere Screeningexperimente scheint er geeignet, da er eine gute Filterleistung aufweist und dennoch einen erheblichen Teil an Aktiven passieren lässt. Nachteil dieses Pharmakophors ist, dass er keinerlei Information über die **Gestalt** der Bindetasche enthält. Daher können Verbindungen den Filter passieren, die eigentlich sterische *clashes* mit dem Protein hervorrufen. Deshalb wurde versucht, die Gestalt der Bindetasche zu approximieren, indem sog. *excluded volumes* auf den Schweratomen der Aminosäuren der Proteinbindetasche platziert wurden. Analog zu den Untersuchungen von Greenidge et al. [Greenidge et al., 1998] wurden als Toleranzen die van-der-Waals-Radien des jeweiligen Elements für ein bestimmtes Schweratom verwendet, wobei diese zuvor mit empirisch bestimmten Skalierungsfaktoren multipliziert wurden. Als Faktoren wurden 25%, 50% und 75% getestet. Ein Faktor von 50% erwies sich als optimal in Bezug auf Erkennung aktiver Inhibitoren und Reduktion der Suchdatenbank. Auf diese Weise wurden 125 *excluded volumes* definiert; es bleibt anzumerken, dass die *excluded volumes* bei einer Datenbanksuche mit dem BEST-Algorithmus nicht als vollkommen undurchdringlich betrachtet werden. CATALYST berechnet eine van-der-Waals-Wechselwirkungsenergie zwischen *excluded volume* und Ligandatom und erhöht die Konformationsenergie um diesen Betrag.

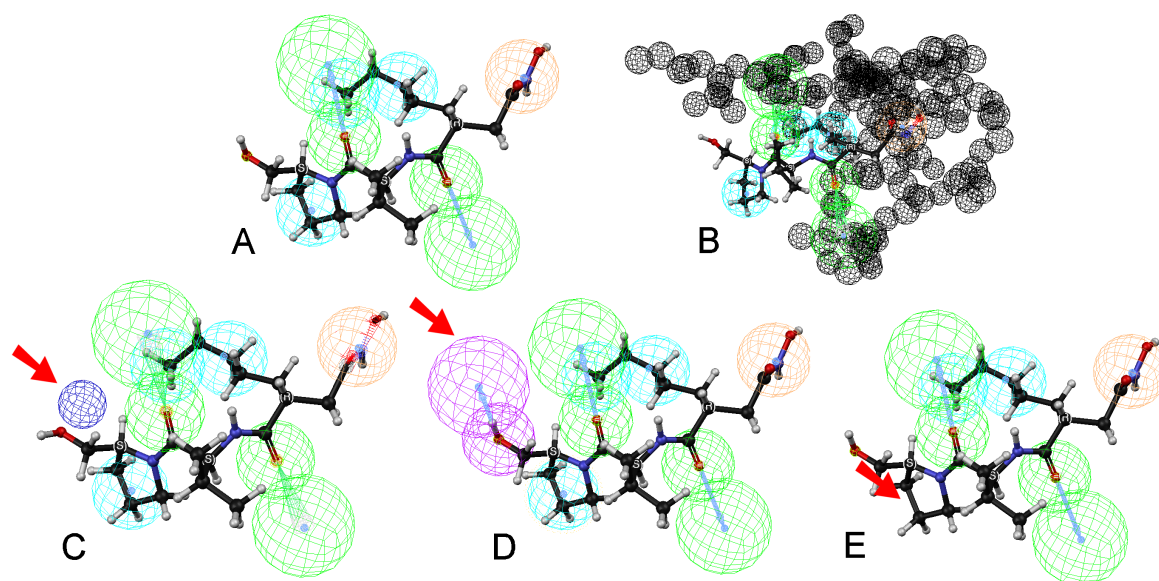


Abbildung 7.7: Darstellung der getesteten Pharmakophorhypothesen. Die roten Pfeile markieren die gegenüber dem Basispharmakophor **A** vorgenommenen Modifikationen. Die in Hypothese **B** eingeführten *excluded volumes* sind auch in den Hypothesen **C–E** enthalten, jedoch aus Gründen der Übersichtlichkeit nicht eingezeichnet. **A:** Basispharmakophor (s. Abb. 7.4); **B:** wie **A**, jedoch mit *excluded volumes*; **C:** wie **B**, jedoch mit zusätzlichem ANY-*feature* (s. Kap. 7.5); **D:** wie **B**, jedoch mit der Donorfunktion des Actinonins; **E:** wie **B**, jedoch ohne hydrophobes *feature* in der S_3' -Tasche. Orange: Metallanker; grün: Akzeptor; violett: Donor; cyan: hydrophob; blau: ANY; schwarz: *excluded volume*.

Pharmakophor **B** weist bezüglich ZINC-COOH und WDI eine etwa 2,5- bis 4,5-fach höhere Spezifität auf als **A**, während nur ein aktives Molekül weniger gefunden wird. Dies rechtfertigt die weitere Verwendung der *excluded volumes*; sie werden daher in den Pharmakophoren **C–E** stets verwendet, auch wenn sie aus Gründen der Übersichtlichkeit in Abb. 7.7 nicht mehr eingezeichnet sind.

Pharmakophor **C** ist ein Versuch, durch Platzierung eines ANY-*features* beim Trp146 der HsPDF1A Selektivitätsaspekte bezüglich der humanen mitochondrialen PDF zu berücksichtigen (s. Kap. 7.5). Keine der aus der Literatur bekannten aktiven Inhibitoren erfüllt diesen Pharmakophor. Geeignete Kandidaten müssten also noch größer als Actinonin sein; Provozierung eines *clashes* mit Trp146, wie es von Fieulaine et al. in [Fieulaine et al., 2005] beschrieben wird, scheint daher keine praktikable Lösung des Selektivitätsproblems zu sein, möchte man bei niedermolekularen Inhibitoren bleiben. Andere substanzielle Unterschiede der Bindetaschen im näheren Umfeld des Me-

tallankers gibt es jedoch nicht, so dass vermutlich andere als pharmakodynamische Unterschiede ausgenutzt werden müssten, um Selektivität für das parasitäre Enzym zu erzielen (s. Kap. 1.1). Denkbar wäre die Ausnutzung von Unterschieden in den ADME-Parametern, wie Verbleiben der Inhibitoren in der Blutzirkulation, unterschiedliche Entgiftungswege, Unterschiede in den Zielorganellen (Mitochondrium vs. Apicoplast) etc.

Pharmakophor **D** unterscheidet sich von **B** durch einen zusätzlichen Donor. Dieser wird im Actinonin von einer Hydroxylgruppe bedient, kein anderer potenter Inhibitor der öffentlich zugänglichen Literatur besitzt dieses *feature*. Da fast keine Verbindung aus den Screening-Datenbanken den Filter passieren, erweist er sich als zu restriktiv, und auf dieses *feature* wird im Folgenden verzichtet.

Wie bereits in Kap. 7.4 ausgeführt, scheint das hydrophobe *feature* in der S_3' -Tasche, welches das Templat Actinonin aufweist, nicht unbedingt erforderlich zu sein. Einerseits ist die Tasche kaum definiert und stark solvensexponiert, andererseits existieren mehrere potente literaturbekannte Inhibitoren, die dieses *feature* nicht besitzen. Daher wurde Pharmakophor **E** getestet, der wie **B** aufgebaut ist, jedoch das hydrophobe *feature* der S_3' -Tasche nicht besitzt. Es wurden sechs zusätzliche Aktive, die nicht durch **B** identifiziert wurden, von **E** gefunden; dabei handelt es sich beispielsweise um Verbindungen wie **58** (s. Tab. 7.2, S. 150) oder **68** (s. Tab. 7.3, S. 151). Der Pharmakophor ist jedoch spezifisch genug, um die ZINC-Datenbank auf insgesamt 467 Moleküle zu reduzieren. Diese Anzahl ist gering genug, um mit zeit- und arbeitsintensiveren Methoden (Docking, visuelle Inspektion etc.) weiter vorzugehen.

7.7 Postprozessierung und Auswahl der Screening-Hits

Ziel dieses virtuellen Screening Projektes war es, ca. 10 bis 15 Verbindungen auszuwählen, die gekauft und anschließend in entsprechenden Assays vermessen werden sollen. Eine genauere Inspektion der 467 Moleküle, die Pharmakophor **E** passieren, zeigt, dass die meisten dieser Verbindungen praktisch keine Ähnlichkeiten zu dem Grundgerüst des peptidischen Actinonin aufweisen. Folglich ist das gewählte Protokoll geeignet, ein sog. *scaffold hopping* [Schneider et al., 1999] durchzuführen, also ausgehend von einer Substanz mit bestimmter biologischer Aktivität ein chemisch nicht verwandtes anderes Grundgerüst zu identifizieren, welches ähnliche biologische Wir-

kungen zeigt.

Um eine Auswahl zu treffen, wurde versucht, möglichst gute Kompromisse zwischen folgenden Faktoren zu erhalten:

- Verfügbarkeit: Nicht alle in der ZINC-Datenbank aufgeführten Verbindungen sind tatsächlich kommerziell erhältlich.
- Chemische Diversität: Nur eines von mehreren chemisch ähnlichen Molekülen wird ausgewählt, um möglichst weiträumig den „sinnvollen“ chemischen Raum abzudecken.
- Visuelle Inspektion des Pharmakophor-Matching: Einige Hits passieren den Filter, erfüllen den Pharmakophor aber eher theoretisch, als dass eine entsprechende Orientierung im Protein wahrscheinlich erscheint.
- FIT-Wert: Moleküle mit hohen FIT-Werten werden bevorzugt; dadurch wird erreicht, dass Strukturen, die den Pharmakophor zwar erfüllen, deren *features* aber an der Grenze der Toleranzsphären zu liegen kommen, aussortiert werden.
- Konformationsenergie: Beim BEST-Fit-Verfahren kann die gemäß dem CATALYST-Kraftfeld errechnete Energie ausgegeben werden. Zwar ist diese nur eine Näherung, wird *in vacuo* berechnet und gibt keinerlei Auskunft über die Protein-Ligand-Wechselwirkungsenergie, dennoch geben außergewöhnlich hohe Werte Hinweise auf sehr ungünstige und verzerrte Geometrien. Werte ab 10 kcal/mol wurden als kritische, ab 15 kcal/mol als sehr unwahrscheinliche Konformationen angesehen.
- Vorhandensein eines Hydroxamatankers: Da Hydroxamate und Formylhydroxylamine die stärksten bekannten Eisenanker darstellen und die Chelatisierung vermutlich einen erheblichen Anteil an der gesamten Bindungsenthalpie hat, sollten Verbindungen mit derartigen Substrukturen bevorzugt werden, selbst wenn andere Faktoren nur suboptimal erfüllt werden.
- Dockinglösungen: Aufgrund des heuristisch-approximativen Charakters der Konformationssuche in CATALYST und der Tatsache, dass bei einer Pharmakophorsuche jede Möglichkeit versucht wird, das erwünschte Wechselwirkungsmuster zu finden, empfiehlt sich ein Docking potentieller Kandidatenmoleküle. Dies erlaubt eine ausgiebigere Konformationssuche und bezieht auch die Wechselwirkungsenergie mit dem Protein ein.

Von den 467 Verbindungen besitzen nur 14 einen Hydroxamatanker. Die meisten Hydroxamate von ihnen waren jedoch entweder nicht käuflich zu erwerben oder zeigten sehr unwahrscheinliche Konformationen und Orientierungen beim Einpassen in den Pharmakophor. Daher wurden auch vier Hydrazide erworben. Die sieben Verbindungen, die schließlich zum Kauf ausgewählt wurden, sind in Tab. 7.5, S. 153 aufgeführt.

Alle aufgelisteten Verbindungen besitzen einen FIT-Wert größer 2.5, was relativ gesehen ein vergleichsweise hoher Wert ist (bei Pharmakophor **E** wäre bei perfektem Fit ein Wert von maximal 5.0 möglich, da fünf *features* vorhanden sind). Dabei nehmen die berechneten Konformationsenergien – abgesehen von **72** – Werte deutlich kleiner 10 kcal/mol an, was darauf hinweist, dass die Moleküle keine extremen Konformationsspannungen erleiden müssen, um den Pharmakophor zu erfüllen. Bei **72** rührt der relativ hohe Wert von 11.75 kcal/mol vermutlich daher, dass die Ethylestergruppe fast senkrecht zur Ebene des benachbarten Thiophenrings stehen muss, um die beiden Akzeptor-*features* gleichzeitig bedienen zu können. Die Unterbrechung des konjugierten π -Elektronensystems kostet einen entsprechenden Energiebetrag.

Im folgenden Abschnitt erfolgt eine genauere Diskussion der sieben ausgewählten Verbindungen im Hinblick auf die Art und Weise, wie sie in das Pharmakophormodell passen und die Untersuchung der Dockinglösungen.

Molekül **70** (Abb. 7.8, links) besitzt ein Bromindol-Grundgerüst, das in eine Spiroverbindung eingebaut ist. Es besitzt nur eine einzige frei rotierbare Bindung und ist daher sehr rigide. Die beiden hydrophoben *features* der S_1' -Tasche werden durch das Bromatom und den Schwerpunkt des Phenylrings bedient. Die Wasserstoffbrückenakzeptor-*features* werden durch eines der Sauerstoffatome des Heterozyklus und das Carbonylsauerstoffatom am Indolsystem abgedeckt.

Die Dockinglösungen (Abb. 7.8, rechts) mit dem Programm GOLD bestätigen diesen Bindemodus zwar; die Orientierung der Lösungen mit den höchsten Scores stimmt mit der pharmakophorbasierten überein. Dennoch sind in den Dockinglösungen die Abstände der Akzeptoren zu den Backbone-NH-Gruppen in allen Fällen für echte Wasserstoffbrücken zu groß (niedrigste gefundene Distanz zwischen der Carbonylgruppe des Ringes und der Ile105-NH-Gruppe: 3.77 Å ; Spiro-O-atom und Gly154-NH: 4.04 Å). Dass der Pharmakophor dennoch erfüllt wurde, liegt an den erlaubten Toleranzen von 1.5 Å.

Verbindung **71** (Abb. 7.9) war nur als Racemat erhältlich. Gemessen an FIT-Wert

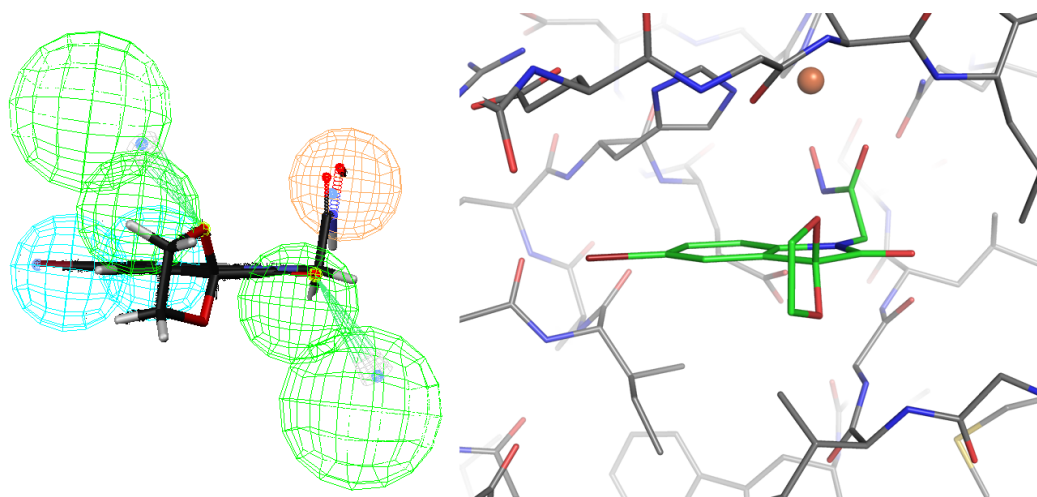


Abbildung 7.8: **Pharmakophor-Matching (links) und Dockinglösung (rechts) der Verbindung 70.** Die Farbkodierung der Pharmakophorabbildung links ist dieselbe wie in Abb. 7.4, *excluded volumes* sind der Übersichtlichkeit wegen nicht dargestellt. In der Dockinglösung rechts sind die Kohlenstoffatome des Proteins (PDB:1RL4) in grau, die des Liganden in grün dargestellt. Das Fe^{2+} -Ion ist als rötliche Kugel zu sehen. Rot: Sauerstoff, blau: Stickstoff, rotbraun: Brom.

und Konformationsenergie scheint das R-Enantiomer (Abb. 7.9, oben) besser in den Pharmakophor zu passen. Allerdings unterscheiden sie sich nicht sonderlich stark in diesen Parametern. Die Zuweisung der Molekülfunktionalitäten auf die Pharmakophor-*features* erfolgt bei beiden Enantiomeren in gleicher Weise (Abb. 7.9, links): Die Methionin-Seitenkette deckt die hydrophoben *features* ab, die Carbonylgruppen des Phthalimid-Rings besetzen die Wasserstoffbrückenakzeptor-*features*.

Aus Sicht der Dockinglösungen scheint der Ligand weniger geeignet: Beim R-Enantiomer (Abb. 7.9, rechts oben) werden nur wenige (vier) Lösungen gefunden, bei denen der Methionin-Rest die S_1' -Tasche besetzt. Bei diesen Lösungen wird von der einen Carbonylgruppe eine Wasserstoffbrücke zum Ile105 (2.75 Å) und eine schwächere zum Gly106 (3.48 Å) ausgebildet, die zweite Carbonylgruppe hingegen ist ungünstigerweise sehr dicht (ca. 2.1 Å) der Carbonylgruppe des Gly154 angenähert. Neben dieser eher selten beobachteten Geometrie sind zwei weitere Bindemoden zu finden, bei denen der Methionin-Rest in die S_2' -Tasche orientiert ist, der Phthalimid-Ring hingegen besetzt die S_1' -Tasche. Bei dem Bindemodus mit günstigerem Score fungiert das Schwefelatom als Akzeptor gegenüber Ile105, die Carbonylsauerstoffatome des Phthalimidringes bil-

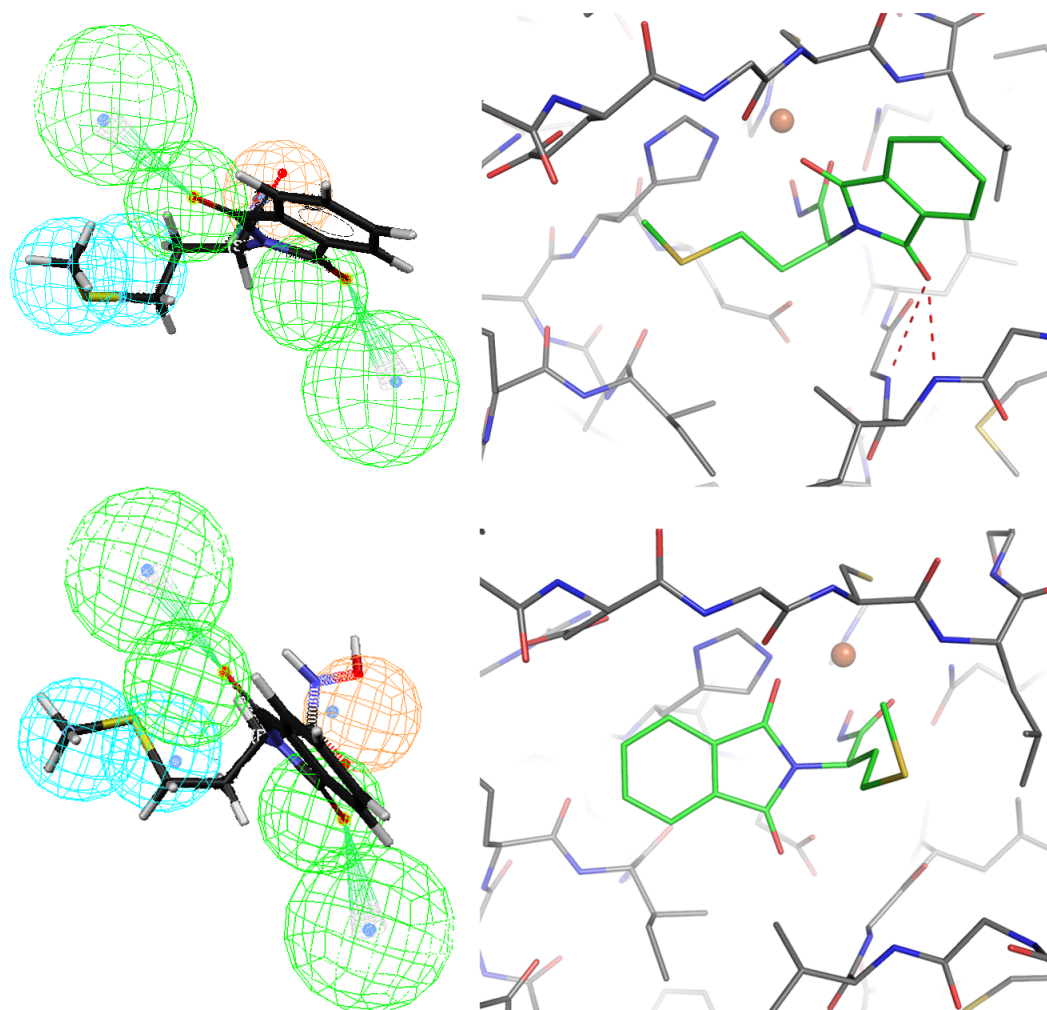


Abbildung 7.9: Pharmakophor-Matching (links) und Dockinglösung (rechts) des R- (oben) und S-Enantiomers (unten) von 71. Die roten, gestrichelten Linien zeigen Wasserstoffbrückenbindungen an. Rot: Sauerstoff, blau: Stickstoff, gelb: Schwefel.

den hingegen keinerlei Wasserstoffbrücken aus. Stattdessen ist die Distanz zu Atomen des Proteins mit ca. 2.50 Å für nichtbindende Kontakte bedenklich niedrig. Bei der schlechter bewerteten Lösung (ohne Abb.) steht der Phthalimidring orthogonal zur günstigeren Lösung; eine der Carbonylgruppen kann dadurch eine schwache H-Brücke zu Gly106 ausbilden (3.54 Å), die andere bildet dafür eine ungünstig nahe Distanz von 1.87 Å zu einer Backbone-Carbonylgruppe aus, so dass diese Dockinglösung noch unwahrscheinlicher erscheint.

Beim S-Enantiomer (Abb. 7.9, rechts unten) besitzen die Dockinglösungen sehr ähnliche Geometrien wie die zuletzt beschriebene Lösung. GOLD findet hier allerdings keine Alternativen.

Bei Verbindung **72** (Abb. 7.10) wird die hydrophobe S_1' -Tasche durch ein Tetrahydrobenzothiophen besetzt. Die beiden Akzeptor-*features* werden durch eine Estergruppe bedient, allerdings liegen die *matching-points* an der Grenze der Toleranzsphären, so dass zu erwarten ist, dass maximal eine Wasserstoffbrücke zum Protein ausgebildet werden kann.

Die Dockinglösungen dieses Liganden waren sehr inhomogen, und nur bei wenigen Lösungen war eine Koordination des Eisen-(II)-Ions vorhanden, allerdings monodentat. Aus diesen Gründen wurde das Docking unter Verwendung eines sog. *scaffold-constraints* wiederholt, wobei die Lage des Hydrazidankers in Analogie zu der des Hydroxamatankers aus Kristallstrukturen vorgegeben wurde. Überraschenderweise ragt bei allen Lösungen die Ethylestergruppe in die S_1' -Tasche; dabei bildet das Alkoxy-Sauerstoffatom einen polaren Kontakt zur NH-Gruppe von Gly154 in einer Distanz etwas jenseits einer Wasserstoffbrücke (3.64 Å) aus. Das Carbonylsauerstoffatom ist mit 2.28 Å ungünstig dicht an dem Carbonylsauerstoffatom desselben Glycins positioniert. Das Schwefelatom im Thiophenring bildet hierbei eine H-Brücke zur NH-Gruppe des Ile105 aus (3.48 Å); allerdings bleibt zu bemerken, dass Thiophen-Schwefelatome i.A. eher schlechte Wasserstoffbrückenakzeptoren sind. Zusammenfassend spricht das Docking hier also eher gegen die Ergebnisse aus der Pharmakophorsuche.

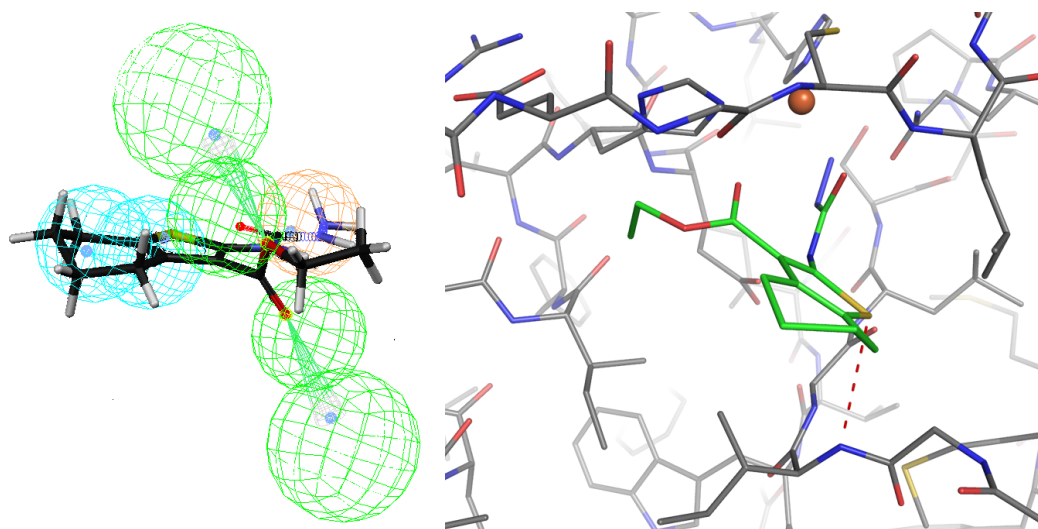


Abbildung 7.10: Pharmakophor-Matching (links) und Dockinglösung (rechts) der Verbindung **72**. Die rote, gestrichelte Linie zeigt eine Wasserstoffbrückenbindung an. Rot: Sauerstoff, blau: Stickstoff, gelb: Schwefel.

Wie bei Molekül **72** werden bei Molekül **73** (Abb. 7.11) die beiden hydrophoben *features* der S_1' -Tasche durch eine Tetrahydrobenzothiophen-Struktur erfüllt. Eine Car-

bonylfunktion bedient den Akzeptor zum Gly154, der zweite Akzeptor zum Ile105 wird dagegen durch eine Aminogruppe bedient.

Die Dockinglösungen bestätigen weitgehend die Möglichkeit eines solchen Bindungsmodus, allerdings werden keine Lösungen mit zwei Wasserstoffbrücken gefunden, sondern nur solche, bei denen das Carbonylsauerstoffatom eine Wasserstoffbrücke zum Gly154 ausbildet.

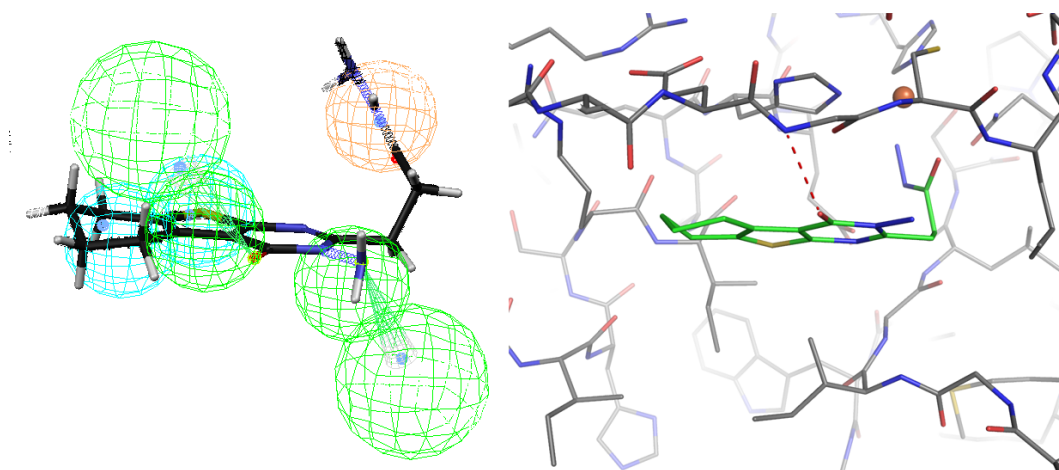


Abbildung 7.11: **Pharmakophor-Matching (links) und Dockinglösung (rechts) der Verbindung 73.** Die rote, gestrichelte Linie zeigt eine Wasserstoffbrückenbindung an. Rot: Sauerstoff, blau: Stickstoff, gelb: Schwefel.

Bei Molekül **74** werden die hydrophoben *features* von einer Methylgruppe und einem Bromatom abgedeckt. Der Akzeptor zum Gly154 wird von der Carbonylgruppe, der zum Ile105 von einem Schwefelatom bedient.

Das Docking bestätigt weitgehend eine solche Orientierung im Protein. Die H-Brücke zum Gly154 wird mit einem Abstand von 2.70 Å ausgebildet. Allerdings wird die zweite Wasserstoffbrücke zwischen Schwefelatom und Ile105 nicht reproduziert.

In Verbindung **75** erfüllen das Zentrum des Phenylrings und eine Methylgruppe die hydrophoben S_1' -*features*. Beide Akzeptorfunktionen werden von den Sauerstoffatomen der Sulfonamidgruppe besetzt. Diese liegen am Rande der jeweiligen Toleranzsphäre, so dass im Protein vermutlich nur eine H-Brücke ausgebildet werden kann.

Hinsichtlich der hydrophoben Wechselwirkung wird die Orientierung im Pharmakophor durch das Docking bestätigt, im Bereich der Sulfonamidgruppierung wird hingegen eine andere Konformation eingenommen: Zur NH-Gruppe des Ile105 bildet ein Sulfonsauerstoffatom eine H-Brücke aus (2.71 Å), das andere bildet eine Wasserstoffbrücke zur

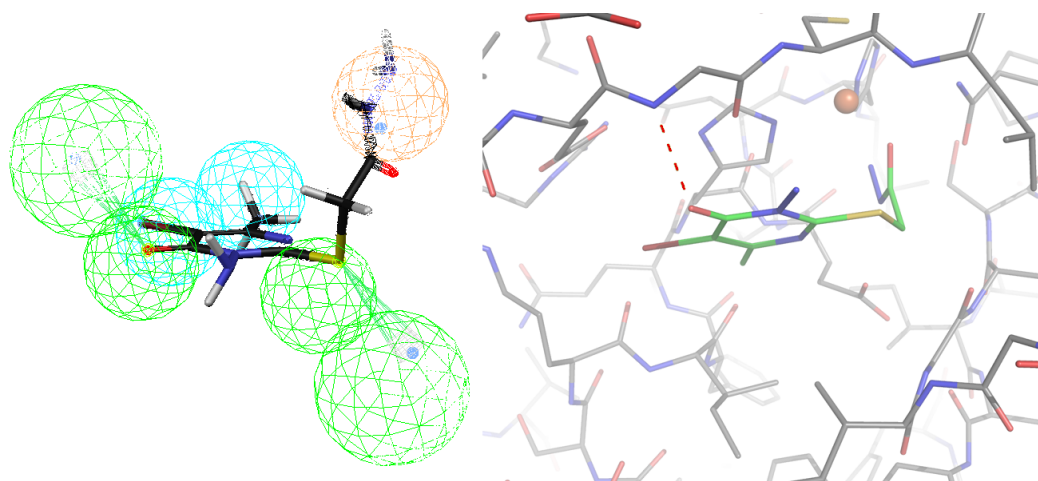


Abbildung 7.12: **Pharmakophor-Matching (links) und Dockinglösung (rechts) der Verbindung 74.** Die rote, gestrichelte Linie zeigt eine Wasserstoffbrückenbindung an. Rot: Sauerstoff, blau: Stickstoff, gelb: Schwefel, rotbraun: Brom.

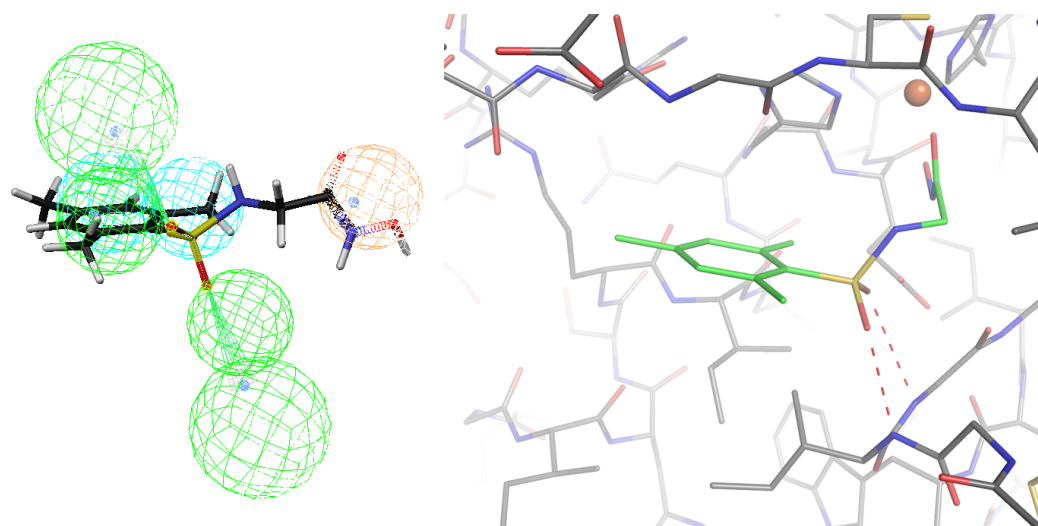


Abbildung 7.13: **Pharmakophor-Matching (links) und Dockinglösung (rechts) der Verbindung 75.** Die roten, gestrichelten Linien zeigen Wasserstoffbrückenbindungen an. Rot: Sauerstoff, blau: Stickstoff, gelb: Schwefel.

NH-Gruppe von Gly106 aus (3.21 Å). Demzufolge schlägt das Docking einen alternativen, aber günstigen Bindemodus vor.

Die beiden hydrophoben *features* werden beim Molekül **76** durch das Zentrum eines Phenylrings und ein Chloratom abgedeckt, die beiden Akzeptoren werden durch die zwei Sauerstoffatome der Sulfonamidgruppe besetzt (Abb. 7.14).

In Bezug auf die hydrophoben *features* wird die Lage des Liganden im Pharmakophor durch das Docking bestätigt, die Sulfonamidgruppe bildet allerdings nur eine H-Brücke zum Ile105 aus (2.91 Å), der 4-Chlorphenylrest zeigt in die entgegengesetzte Richtung. Der Hydrazidanker besitzt zwar einen geeigneten Abstand zum Eisenion, jedoch ist die Direktionalität für eine Koordination ungeeignet.

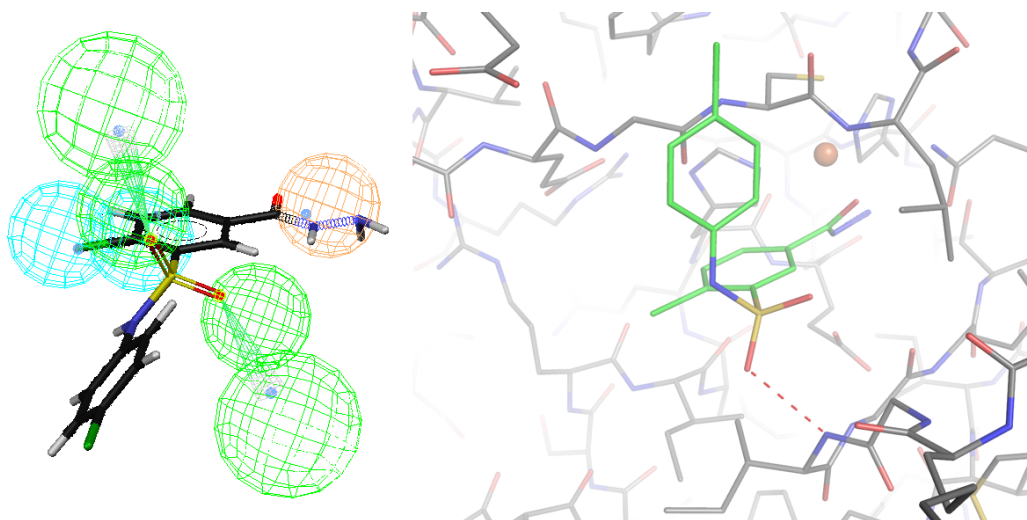


Abbildung 7.14: **Pharmakophor-Matching (links) und Dockinglösung (rechts) der Verbindung 76.** Die rote, gestrichelte Linie zeigt eine Wasserstoffbrückenbindung an. Rot: Sauerstoff, blau: Stickstoff, gelb: Schwefel, grün: Chlor.

Zusammenfassend kann festgehalten werden, dass als vorrangiges Kriterium für die Selektion zum Kauf der Verbindungen die Erfüllung des Pharmakophors verwendet wurde. Bei den Verbindungen **71** und **72** konnten keine sinnvollen Dockingposen ohne ungünstige Kontakte gefunden werden; hier stehen die Dockinggeometrien den Pharmakophorergebnissen eindeutig entgegen. In den anderen Fällen werden sinnvolle Lösungen gefunden, denen zufolge meist nicht alle *features* wie beim Pharmakophor-matching gefunden werden. Insbesondere war zu beobachten, dass die beiden recht nahe gelegenen Wasserstoffbrückenakzeptoren im Pharmakophormatching grenzwertig durch eine funktionelle Gruppe (Ester, Sulfon) simultan bedient werden konnten. Dies hängt allerdings mit den Toleranzsphären zusammen. Unabhängig von der Güte der Dockinglösungen scheinen solche Konstellationen allerdings ausgeschlossen. Dennoch sind die erhaltenen Lösungen an sich sinnvoll und meist mit der Orientierung des Pharmakophormatching kompatibel. Inwieweit die Auslassung einer Wasserstoffbrücke zu einem drastischen Affinitätsverlust führen wird, ist schwer abzuschätzen.

7.8 Identifizierung weiterer potentieller Hydroxamatliganden

Da nur äußerst wenige Moleküle mit Hydroxamatanker den Pharmakophor in geeigneter Weise erfüllten und trotzdem einige weitere Substanzen bestellt werden sollten, war es als nächstes das Ziel, Hydroxamate zu finden, die wenigstens einige günstige Wechselwirkungen mit der PDF eingehen können. Dazu wurden zunächst sämtliche Hydroxamate aus dem „*druglike subset*“ der ZINC-Datenbank extrahiert; dies sind insgesamt nur 434 Verbindungen. Diese wurden mit dem Programm FlexX in die Bindetasche der PfPDF gedockt (PDB-code: 1RL4), wobei die Geometrie des Hydroxamatankers vorgegeben wurde. Sie wurde vom Actinoninligand der Kristallstruktur 1G2A abgeleitet. Unter diesen Bedingungen konnten nur für 272 Moleküle Lösungen erzeugt werden.

Die Dockinglösungen wurden visuell inspiziert im Hinblick auf ihre Plausibilität und daraufhin, ob sie einige der Pharmakophor-*features* aufweisen. Auf diese Weise wurden vier weitere Moleküle ausgewählt, sie sind in Tab. 7.6, S. 154 gelistet.

Im folgenden Abschnitt soll kurz auf die Plausibilität der entsprechenden Dockinglösungen und günstige Wechselwirkungen mit dem PDF-Protein eingegangen werden.

Eine Besonderheit von Verbindung **77** ist das β -Dicarbonylsystem; daher liegt wahrscheinlich ein signifikanter Anteil der Verbindung als Enol vor. Bei der Dockinglösung mit dem besten Score (Abb. 7.15, links oben) werden zwei Wasserstoffbrückenbindungen jeweils zur NH-Gruppe von Ile105 (3.08 Å) und Gly106 (3.43 Å) ausgebildet. Gleichzeitig weist die α -Carbonylgruppe einen Abstand von 2.50 Å zur Carbonylgruppe von Gly106 auf. Wenn die α -CO-Gruppe enolisiert vorliegt, so wäre dieser Kontakt als weitere Wasserstoffbrücke denkbar. Nachteilig an der gedockten Geometrie ist, dass die S_1' -Tasche vollkommen unbesetzt bleibt; es ist jedoch keine Kristallstruktur bekannt, in der die S_1' -Tasche unausgefüllt bleibt.

Die Dockinggeometrie von Molekül **78** erfüllt zwei *features* (Abb. 7.15, rechts oben): Einer der Imid-Sauerstoffatome bildet eine Wasserstoffbrücke zu Gly154 aus; gleichzeitig ist die Pentylkette am Hydroxamatanker so positioniert, dass sie wenigstens teilweise die hydrophoben *features* der S_1' -Tasche abdeckt.

Die Dockinglösung von Struktur **79** zeigt eine Wasserstoffbrücke zur NH-Gruppe von Ile105 (3.21 Å) (Abb. 7.15, links unten). Der Abstand der zweiten Phthalimid-

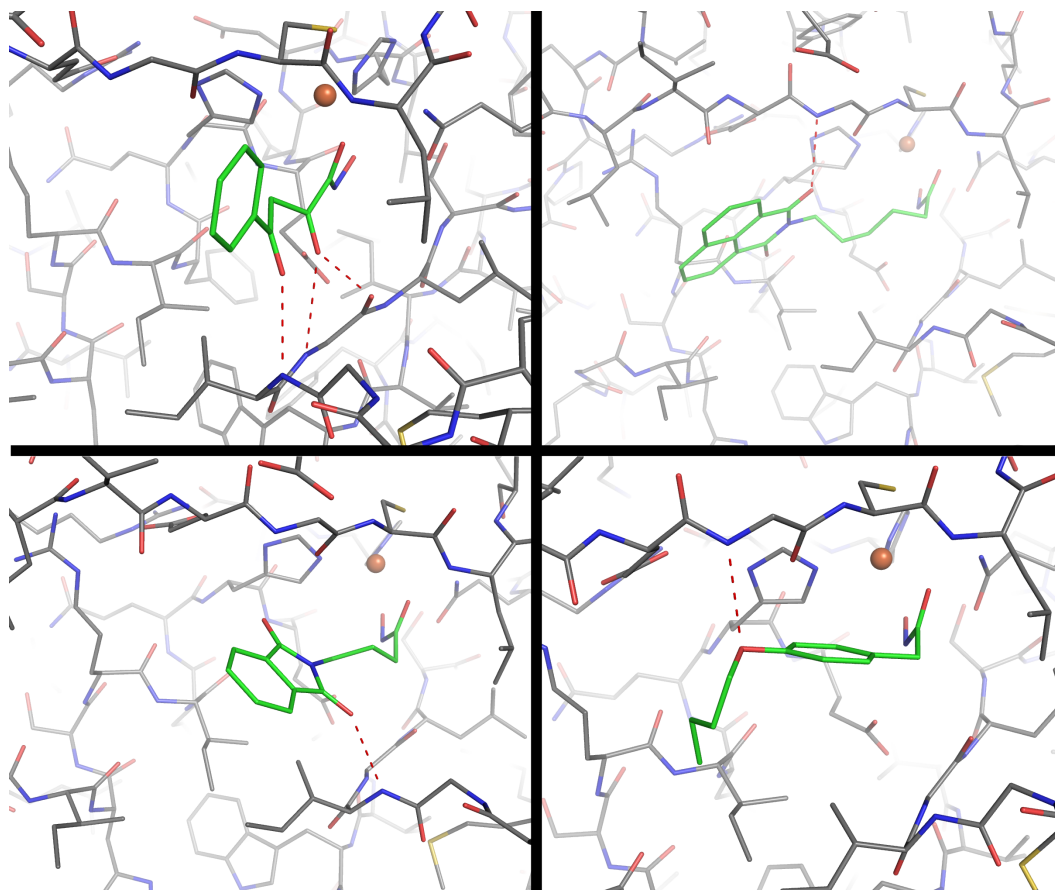


Abbildung 7.15: FlexX-Dockinglösungen der Moleküle **77** bis **80**. Links oben: **77**, rechts oben: **78**, links unten: **79**, rechts unten: **80**. Rote, gestrichelte Linien zeigen Wasserstoffbrückenbindungen an. Grau: Kohlenstoff des Proteins (PDB-code: 1RL4), blau: Stickstoff, grün: Kohlenstoff des Liganden, rot: Sauerstoff, orange Kugel: Fe^{2+} -Ion.

Carbonylgruppe zur Aminogruppe von Gly154 ist mit 3.77 Å etwas zu groß. Die S_1' -Tasche bleibt jedoch unbesetzt.

Für Molekül **80** (Abb. 7.15, rechts unten) findet FlexX sehr einheitliche Dockinglösungen, Unterschiede treten nur in der Konformation der Butylkette auf. Das Arylethersauerstoffatom bildet eine schwache Wasserstoffbrücke zum Gly154 aus (3.16Å). Der Phenylring besetzt einen Teil der hydrophoben S_1' -Tasche.

7.9 Ergebnisse und Diskussion

In diesem Abschnitt wurden elf Verbindungen selektiert, die kommerziell erhältlich sind und eine erhöhte Wahrscheinlichkeit besitzen sollten, Peptiddeformylasen zu inhibieren. Das initiale Ziel ist die Inhibierung der PDF1B aus *Plasmodium falciparum*. Das entwickelte Pharmakophormodell sollte jedoch auch auf die Erfordernisse für die Ligandbindung bei anderen PDFs übertragbar sein, insbesondere auf diejenigen der nahe verwandten PDF1Bs in pflanzlichen Chloroplasten und in gramnegativen Bakterien. Dies liegt in der hohen Konservierung der Bindetasche dieser Enzyme begründet. Ein Vergleich der spärlichen quantitativen Inhibitionsdaten von PfPDF mit denen bakterieller PDFs lässt allerdings vermuten, dass die identifizierten Verbindungen die PfPDF deutlich schwächer hemmen werden als die PDF1Bs gramnegativer Bakterien: Es konnten nur zwei IC_{50} -Werte für die PfPDF in der Literatur gefunden werden. Ein Thiol-Inhibitor hemmt PfPDF mit einem IC_{50} von $57\ \mu\text{M}$, die PDF von *E. coli* (eisenhaltig) jedoch mit einer K_d von $19\ \text{nM}$. Und für das hochpotente Actinonin ist der IC_{50} gegen PfPDF $2.5\ \mu\text{M}$, während der K_d -Wert von der *E. coli*-PDF bei $3\ \text{nM}$ liegt. Inhibitoren der PDF scheinen demnach – soweit man das aus diesen zwei Daten schließen kann – tendenziell das *P. falciparum*-Enzym deutlich schwächer zu hemmen. Strukturelle Gründe dafür werden in der Referenz [Kumar et al., 2002] diskutiert.

Die Auswahl der Verbindungen stützt sich in erster Linie auf das Pharmakophorscreening. Das Docking bestätigt in vielen Fällen in etwa die Ergebnisse des Pharmakophormatchings, allerdings werden selten Lösungen gefunden, bei denen tatsächlich alle *features* erfüllt werden. In wenigen Fällen werden überhaupt keine sinnvollen Dockinglösungen gefunden. Es war leider festzustellen, dass es nur verhältnismäßig wenige *druglike* Substanzen gab, die tatsächlich käuflich sind, geeignete Metallanker besitzen und zusätzlich die Screening-Filter zufriedenstellend erfüllen. Noch schwieriger wird es, wenn man Hydroxamate bevorzugt, die potentesten bekannten Fe^{2+} -Anker. Insofern bleibt zu berücksichtigen, dass die hier vorgeschlagenen Substanzen allesamt Kompromisse darstellen und keine von ihnen aufgrund der durchgeführten Modellierungsstudien als Idealkandidat anzusehen wäre.

Die Substanzen werden auf PDF-Inhibition in der Gruppe von Prof. Dr. THIERRY MEINNEL am CNRS in Frankreich vermessen. Zum Zeitpunkt der Fertigstellung der vorliegenden Arbeit lagen erste Messergebnisse an der PDF1B aus *E. coli* vor. Die vollständigen Daten sind bisher nicht zugänglich, jedoch wurde für Verbindung **70** ein

IC₅₀ von **60 nM** gemessen, bei **78** ergab sich ein IC₅₀ von **190 nM** (persönliche Mitteilung von Prof. Dr. T. MEINNEL). Interessanterweise stammen die Moleküle jeweils aus **unterschiedlichen Screening-Ansätzen**: Molekül **70** wurde über die reine Pharmakophorsuche gefunden, gemäß den Dockinglösungen hingegen wäre es nicht in der Lage, Wasserstoffbrücken auszubilden und wäre demnach nicht selektiert worden. Verbindung **78** hingegen wurde nur durch das Docking entdeckt, demzufolge das Molekül einige, nicht aber alle *features* des Pharmakophors erfüllen kann. Dies unterstreicht den **komplementären Charakter** der beiden angewandten VS-Techniken. Eine parallele Anwendung kann zumindest teilweise helfen, die Unzulänglichkeiten und Approximationen der einzelnen Ansätze zu kompensieren.

Eine weitere Besonderheit der Verbindung **78** ist, dass sie unter dem Namen **Scriptaid**, Scriptide oder GCK 1026 als potenter **Histondeacetylase-Inhibitor** bekannt ist [Su et al., 2000]. Histondeacetylasen (HDACs) katalysieren die Deacetylierung von LysNε-Ketten in Histonen. Dadurch wird die Struktur des Chromatins modifiziert, so dass die DNA fester an die Histone gebunden wird. Dadurch führt die Deacetylierung zur Verringerung der Expressionsrate zahlreicher Gene. Die Inhibition von HDACs bewirkt eine Hemmung der Zellproliferation und induziert die Differenzierung oder Apoptose von Tumorzellen. Daher werden Inhibitoren der HDACs als interessante potentielle Kandidaten in der Krebstherapie angesehen. HDACs enthalten im aktiven Zentrum ein Zn²⁺-Ion, an das höchstwahrscheinlich der Hydroxamatanker des Scriptaid bindet.

Demnach ist anzunehmen, dass Scriptaid einen **dualen Wirkungsmechanismus** besitzen könnte. Dies ist bei der Interpretation eines *in vitro*-Assays gegen *P. falciparum* neben der gefundenen PDF-Inhibition als mögliche Wirkursache zu berücksichtigen. Kürzlich wurden Scriptaid und verschiedene andere HDAC-Inhibitoren als potente Inhibitoren des Wachstums von *Toxoplasma gondii*, einem anderen Vertreter der *Apicomplexa* und Erreger der Toxoplasmose, in *in vitro*-Kulturen nachgewiesen [Strobl et al., 2007].

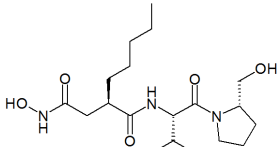
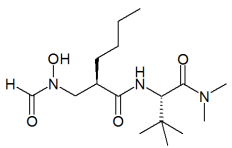
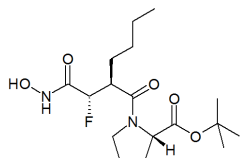
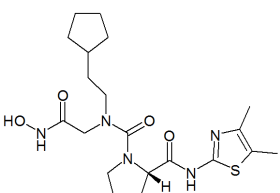
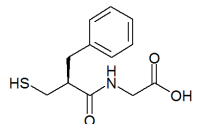
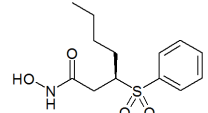
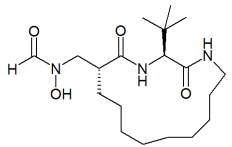
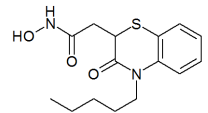
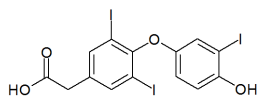
Ferner kann es lohnenswert sein, im Bereich bekannter HDAC-Inhibitoren fokussiert nach PDF-Inhibitoren zu suchen. Die HDAC-Inhibition muss auch als Quelle unerwünschter Nebenwirkungen bei einer Anwendung am Menschen von Scriptaid-ähnlichen Verbindungen berücksichtigt werden.

Über die übrigen Substanzen aus dem Screening liegen noch keine Informationen vor, ob sich noch weitere Hits mit mikromolarer Inhibition darunter befinden. Folgende

Messungen sind geplant:

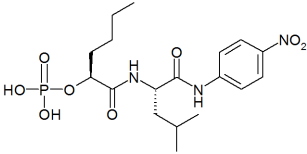
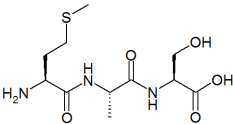
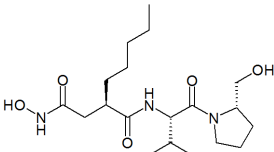
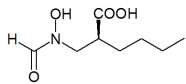
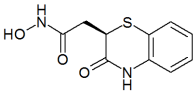
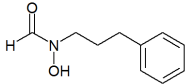
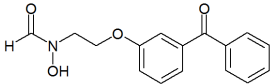
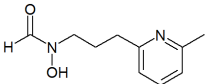
- Ein Ganzzellassay an mit *P. falciparum* infizierten Erythrozytenkulturen. Der Vorteil dieses Tests ist die direkte Ablesbarkeit der Effekte auf den Parasiten. Der Nachteil besteht darin, dass ein Absterben der Parasiten durch andere toxische Effekte (z. B. HDAC-Inhibition) als durch PDF-Inhibition verursacht worden sein kann; im Gegenzug kann Wirkungslosigkeit der Substanzen auch auf Transport- oder Metabolismusphänomene o.ä. zurückzuführen sein, auch bei vorhandener PDF-Inhibition.
- Enzymassays gegen verschiedene PDFs. Endziel soll ein Test gegen die PDF1B von *P. falciparum* sein. Zur Zeit stehen unseren Kooperationspartnern allerdings nur Surrogatenzyme aus *E. coli* (PDF1B) und *Arabidopsis thaliana* (PDF1A und 1B) zur Verfügung. Dennoch ist die Inhibition bakterieller und pflanzlicher PDFs auch als solche interessant (Antibiotika, Herbizide).
- Da mindestens zwei Substanzen die *E. coli*-PDF inhibieren, sollen auch minimale Hemmkonzentrationen (MIC, *Minimal Inhibitory Concentration*) des sichtbaren Bakterienwachstums als Maß für die antibiotische Potenz bestimmt werden.

Tabelle 7.2: Beispiele von PDF-Inhibitoren verschiedener chemischer Strukturklassen mit Affinitätsdaten. (Daten aus [Boularot et al., 2004])

Nr.	Struktur	Bezeichnung	Affinitätswert(e)
53		Actinonin	$K_d(Ec\ Fe) = 3\text{ nM}$ $IC_{50}(Pf\ Fe) = 2.5\ \mu\text{M}$ $MIC(S.\ aureus) = 8\text{--}16\ \mu\text{g/ml}$
54		BB-3497	$K_d(Ec\ Ni) = 7\text{ nM}$ $MIC(S.\ aureus) = 4\text{--}16\ \mu\text{g/ml}$
55		VRC-4071	$IC_{50}(Ec\ Ni) = 2\text{ nM}$ $MIC(S.\ aureus) = 2\text{--}8\ \mu\text{g/ml}$
56		VRC-4307	$IC_{50}(Ec\ Ni) = 2\text{ nM}$ $MIC(S.\ pneumonia) = 1\ \mu\text{g/ml}$
57		Thiorphan	$K_d(Ec\ Ni) = 52\ \mu\text{M}$
58		Ro-66-0376	$IC_{50}(Ec\ Fe) = 16\text{ nM}$ $MIC(S.\ aureus) = 128\ \mu\text{g/ml}$
59		—	$K_d(Ec\ Co) = 0.7\text{ nM}$ $MIC(E.\ coli) = 12\ \mu\text{g/ml}$
60		—	$IC_{50}(Ec\ Fe) < 5\text{ nM}$
61		—	$K_d(Ec\ Ni) = 2.8\ \mu\text{M}$ $MIC(E.\ coli) = 64\ \mu\text{g/ml}$

Ec: *E. coli*, *Pf*: *P. falciparum*, MIC: Minimal Inhibitory Concentration.

Tabelle 7.3: Auflistung der Kristallstrukturen, die zur Pharmakophorableitung herangezogen wurden.

Nr. ^{a)}	Ligand	Name	PDB-code	Organismus	Auflösung [Å]
62		—	1BSJ	<i>E. coli</i>	2.5
63		MAS-Substrat	1BS8	<i>E. coli</i>	2.2
64		Actinonin	1IX1	<i>P. aeruginosa</i>	1.9
64	s.o.	Actinonin	1LQY	<i>B. stearoothermophilus</i>	1.9
64	s.o.	Actinonin	1LRY	<i>P. aeruginosa</i>	2.6
64	s.o.	Actinonin	1Q1Y	<i>S. aureus</i>	1.9
64	s.o.	Actinonin	1WS1	<i>B. cereus</i>	2.0
65		—	1RL4	<i>P. falciparum</i> (3D7)	2.2
66		—	1S17	<i>P. aeruginosa</i>	2.0
67		SB485345	2AI7	<i>S. pneumoniae</i>	2.0
67	s.o.	SB485345	2AI8	<i>E. coli</i>	1.7
68		SB543668	2AIA	<i>S. pneumoniae</i>	1.7
69		SB505684	2AIE	<i>S. pneumoniae</i>	1.7

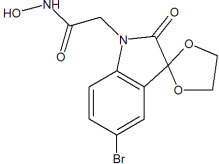
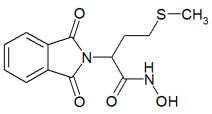
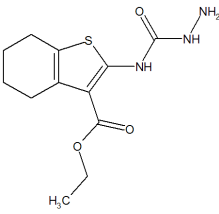
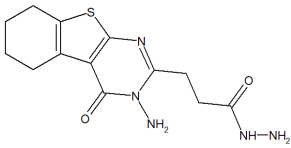
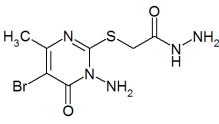
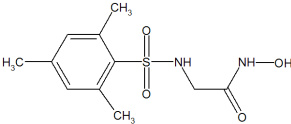
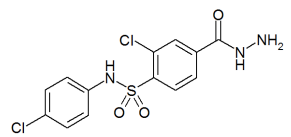
^{a)}Die Nummerierung bezieht sich ausschließlich auf die Liganden.

Tabelle 7.4: Ergebnisse der verschiedenen Pharmakophorhypothesen aus Abb. 7.7.

Es sind jeweils die absoluten und in Klammern die relativen Zahlen (in %) der die entsprechende Hypothese passierenden Moleküle angegeben.

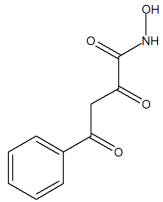
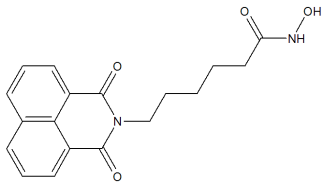
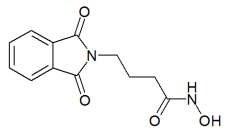
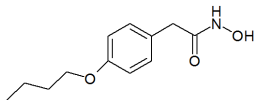
Hypothese	Aktive	ZINC-COOH	ZINC-noCOOH	WDI
A	13 (39.4)	31 (0.041)	203 (3.9)	531 (0.79)
B	12 (36.4)	27 (0.035)	75 (1.4)	117 (0.17)
C	0 (0.0)	0 (0.0)	7 (0.13)	34 (0.05)
D	1 (3.0)	2 (0.0026)	0 (0.0)	10 (0.015)
E	18 (54.5)	59 (0.077)	408 (7.8)	342 (0.51)
Gesamt (100%)	33	76275	5223	67050

Tabelle 7.5: Liste der selektierten Hitstrukturen aus ZINC-noCOOH, die durch Pharmakophor E identifiziert wurden.

Nr.	Struktur	Code	FIT-Wert	Konforma- tionsenergie [kcal/mol]	Anbieter
70*		AG-205- /36450018	2.80	1.22	Specs
71		A3771- /0159975	S: 2.91; R: 3.24	S: 4.46; R: 2.21	ChemicalBlock
72		6049378	3.55	11.75	ChemBridge
73		NS19915	3.39	2.32	Nanosyn
74		6518	3.39	6.10	Toslab
75		NS42048	3.17	3.59	Nanosyn
76		EN300- 03762	3.33	5.55	Enamine

*: aktiv; $IC_{50}(\text{EcPDF1B}) = 60 \text{ nM}$

Tabelle 7.6: Auflistung der vier durch FlexX-Docking identifizierten Moleküle.

Nr.	Struktur	Code	Anbieter
77		BB_SC-3103	InterBioScreen
78*		5217444	ChemBridge
79		A3685/0156331	Chemical Block
80		ka-00853	Aurora Feinchemie

*: aktiv; $IC_{50}(\text{EcPDF1B}) = 190 \text{ nM}$

8 Anhang

8.1 Verwendete Hardware

Sämtliche Berechnungen dieser Arbeit wurden auf einem Linux PC (Intel Pentium 4, 3.0 GHz, 1 GByte RAM) durchgeführt, mit Ausnahme der GOLPE-Berechnungen. Diese wurden auf einer Silicon Graphics O2 Workstation mit einem 225 MHz MIPS R10000 Prozessor mit 512 MByte RAM ausgeführt.

8.2 Verwendete Software

Für Aufbau, Modifikation und Ladungsberechnung von Molekülen sowie zur Durchführung der 3D-QSAR-Studien wurde das Programmpaket SYBYL (SYBYL 7.1 bis 7.3) verwendet. Die Automatisierung zahlreicher Prozeduren wurde in der SYBYL *Programming Language* (SPL) implementiert.

Die Konvertierung von 2D- zu 3D-Koordinaten erfolgte mit dem Konformationsgenerator CORINA (CORINA 3.1).

Das manuelle Alignment der Moleküle für die 3D-QSAR-Studien in den Proteinbindetaschen sowie die nachfolgende Minimierung mit dem MAB-Kraftfeld wurden mit dem Programm MOLOC (MOLOC, 2006) durchgeführt.

Das automatische Alignment wurde mit Hilfe der Python-Schnittstelle *pyflexs* des Programms FlexS (FlexS 1.20.2) realisiert.

Zur Berechnung der MACCS-Keys, der VSA-Deskriptoren und zur automatischen Ableitung eines optimalen VSA-QSAR-Modells wurde das Programmpaket MOE (MOE 2006) verwendet.

Die Berechnung von Interaktionsfeldern wurde mit dem Programm GRID (GRID 22a) durchgeführt. Für die chemometrischen Berechnungen wurde das Programmpaket GOLPE (GOLPE 4.5) genutzt.

Die AFMoC-Untersuchungen wurden mit DrugScore (DrugScore 1.2, Optionen ATRAIN und ATEST) und einer *in-house*-Implementierung des SAMPLS-Algorithmus von H. GOHLKE ausgeführt.

Zur Durchführung der Pharmakophor-Generierung und -Datenbanksuchen wurde das Programmpaket CATALYST (CATALYST 4.11) verwendet.

Zum Docking von Liganden in Proteinbindetaschen wurden die Programme FlexX (FlexX 2.0.3) und GOLD (GOLD 3.1.1) benutzt.

Zur Erstellung dreidimensionaler Abbildungen wurde PyMOL (PyMOL 0.99) verwendet.

Zweidimensionale Plots von Graphen zur Datenvisualisierung wurden mit GNUPLOT (GNUPLOT 4.0) erstellt.

8.3 Statistische Parameter der PLS-Analyse

Zur Beschreibung der Qualität von PLS-Modellen existieren verschiedene Maßzahlen. Um Mehrdeutigkeiten zu vermeiden, werden diese statistischen Parameter im Folgenden kurz beschrieben.

Um die optimale Anzahl a an PLS-Komponenten zu bestimmen, wird i.A. eine sog. *Kreuzvalidierung* durchgeführt: Besteht der gesamte Trainingsdatensatz aus n Verbindungen, so wird dieser in eine frei wählbare Zahl (annähernd) gleich großer Gruppen zufällig aufgeteilt. Dann wird ein PLS-Modell unter Auslassung einer dieser Gruppen abgeleitet, und mit dem so erstellten Modell werden die Affinitäten bzw. Selektivitäten der Verbindungen der ausgelassenen Gruppe vorhergesagt. Dieser Prozess wird so oft wiederholt, bis jede Gruppe einmal ausgelassen wurde. Je nach Anzahl der Gruppen spricht man beispielsweise von *leave-one-out* (Anzahl der Gruppen gleich n , jeweils eine Verbindung wird ausgelassen), *leave-10%-out* (zehn Gruppen werden gebildet) etc.

Auf diese Weise erhält man für jede Verbindung einen vorhergesagten pK_i -Wert, $pK_{i,pred}$. Mit $pK_{i,exp}$ wird der entsprechende experimentelle pK_i -Wert bezeichnet.

Die Summe der quadratischen Abweichungen nennt man *PRESS* (*predictive residual sum of squares*):

$$PRESS = \sum_{k=1}^n (pK_{i,pred}(k) - pK_{i,exp}(k))^2 \quad (8.1)$$

Die Summe der quadratischen Abweichungen der experimentellen Werte vom Mittelwert aller pK_i -Werte, $pK_{i,mean}$, wird als SS (*sum of squares*) bezeichnet:

$$SS = \sum_{k=1}^n (pK_{i,exp}(k) - pK_{i,mean}(k))^2 \quad (8.2)$$

Aus dem Quotienten beider lässt sich somit der q^2 -Wert (auch kreuzvalidierter r^2 oder r_{cv}^2) berechnen:

$$q^2 = 1 - \frac{PRESS}{SS} \quad (8.3)$$

Bei einem perfekten QSAR-Modell wäre $PRESS = 0$, so dass q^2 den Wert 1 annehmen würde. Sind $PRESS$ und SS gleich groß, dann nimmt q^2 den Wert 0 an. In diesem Fall wäre das Modell nicht besser als eines, welches für jeden $pK_{i,pred}(k)$ den Mittelwert annimmt. Negative q^2 -Werte sind auch möglich und bedeuten ein Modell, welches schlechter als Raten des Mittelwertes ist. Generell werden Modelle ab einem q^2 von 0.3 als signifikant, ab 0.5 als gut bewertet.

Als entsprechendes Fehlermaß wird s_{PRESS} (*standard deviation of the error of predictions*) definiert:

$$s_{PRESS} = \sqrt{\frac{PRESS}{n - a - 1}} \quad (8.4)$$

Durch Einbeziehung der Anzahl an PLS-Komponenten a werden Modelle mit einer höheren Komplexität bestraft, da der s_{PRESS} -Wert bei solchen Modellen größer ist. Einige Autoren und das Programmpaket GOLPE bevorzugen den $SDEP$ -Wert (*standard error of prediction*), der allerdings nicht die Anzahl an Komponenten berücksichtigt:

$$SDEP = \sqrt{\frac{PRESS}{n}} \quad (8.5)$$

Die optimale Anzahl an Komponenten wird häufig entweder bei maximalem q^2 oder minimalem s_{PRESS} gewählt. Oft führen beide Kriterien zum gleichen Ergebnis. Da jedoch manchmal das Maximum des q^2 -Wertes erst bei sehr hohen Komponentenzahlen ohne signifikante Zunahme des q^2 -Wertes erreicht wird, ist es ratsam, diejenige Anzahl an PLS-Komponenten zu wählen, bei der der q^2 -Wert um nicht mehr als einen bestimmten Anteil (i.d.R. 5%) zunimmt.

Nach der Bestimmung der optimalen Anzahl an PLS-Komponenten wird zur Ableitung des endgültigen Modells eine nicht-kreuzvalidierte PLS-Analyse vorgenommen. Es ergeben sich analog definierte Güte- und Fehlerparameter, wobei hier anstelle des mit Kreuzvalidierung erhaltenen $pK_{i,pred}(k)$ für das k -te Molekül der $pK_{i,fit}(k)$ benutzt wird. Die quadratische Summe der Fehler wird hierbei mit RSS (*residual sum*

of squares) bezeichnet und ist dem *PRESS* analog:

$$RSS = \sum_{k=1}^n (pK_{i,fit}(k) - pK_{i,exp}(k))^2 \quad (8.6)$$

Der quadrierte Korrelationskoeffizient r^2 ist das dem q^2 analoge Korrelationsmaß:

$$r^2 = 1 - \frac{RSS}{SS} \quad (8.7)$$

Als Fehlermaße werden – dem s_{PRESS} und dem $SDEP$ entsprechend – S und $SDEC$ (*standard deviation of error calculation*) definiert:

$$S = \sqrt{\frac{RSS}{n - a - 1}} \quad (8.8)$$

$$SDEC = \sqrt{\frac{RSS}{n}} \quad (8.9)$$

Schließlich gibt der Fischer-Wert F Auskunft über die statistische Signifikanz der gesamten QSAR-Gleichung:

$$F = \frac{r^2}{1 - r^2} \frac{n - a - 1}{a} \quad (8.10)$$

Ist dieser Wert für ein gegebenes Modell größer als ein Grenzwert (den man F-Werttabellen entnehmen kann, abhängig von n , a und einer gegebenen Irrtumswahrscheinlichkeit α), so kann die Nullhypothese, dass alle Regressionskoeffizienten null sind, verworfen werden; in diesem Fall ist das Gesamtmodell statistisch signifikant.

Der aussagekräftigste Parameter wird berechnet, indem **externe** Moleküle mit dem Modell vorhergesagt werden. Bezeichnet man den vorhergesagten pK_i -Wert eines externen Moleküls aus dem Testdatensatz mit $pK_{i,epred}$, so ergibt sich der *predictive* r^2 -Wert als:

$$r_{pred}^2 = 1 - \frac{\sum_{k=1}^n (pK_{i,epred}(k) - pK_{i,exp}(k))^2}{SS} \quad (8.11)$$

9 Zusammenfassung, Summary

9.1 Zusammenfassung

In der modernen Arzneistoffentwicklung unterscheidet man die Phasen der Leitstrukturfindung und der Leitstrukturoptimierung. Die vorliegende Dissertationsschrift beinhaltet Beiträge zu beiden Bereichen.

Der erste Teil der Arbeit befasst sich mit der Entwicklung und Evaluierung von Computernmodellen zur Vorhersage von Affinität und Selektivität und entstammt daher dem Bereich der Leitstrukturoptimierung. Selektivitätsaspekte spielen eine wichtige Rolle, da sie das Risiko von Nebenwirkungen maßgeblich beeinflussen. Zur Modellierung und Vorhersage von Affinitäts- und Selektivitätsparametern wurden QSAR-Methoden angewendet. Das Modellsystem stellten Carboanhydrasen (CAs) dar; diese zinkhaltigen Hydrolasen katalysieren die reversible Hydratisierung von Kohlendioxid zu Bicarbonat und einem Proton. Sie sind daher in eine Vielzahl (patho)physiologischer Prozesse involviert und stellen interessante therapeutische Targets dar. Die zahlreichen Isoenzyme der CAs besitzen im Bereich der Bindetasche hohe Ähnlichkeiten in Bezug auf physikochemische Eigenschaften, so dass die Entwicklung selektiver Inhibitoren kein triviales Problem darstellt.

Im Mittelpunkt der Untersuchungen standen insbesondere 3D-QSAR-Verfahren. Es wurden statistisch hochsignifikante und robuste Modelle abgeleitet, um Affinität und Selektivität von Sulfonamidinhibitoren bezüglich der Isoenzyme CA I, II und IV vorherzusagen. Es zeigte sich, dass die geringen Unterschiede im strukturbasierten Alignment unter Verwendung der drei Bindetaschen nur geringen Einfluss auf die statistischen Parameter besitzen und dass bessere Ergebnisse erzielt werden, wenn für alle Isoenzyme das auf CA II basierende Alignment benutzt wird anstelle des Alignments in der jeweiligen Bindetasche. Ursache hierfür ist wahrscheinlich die Vielzahl an Kristallstrukturen, die für CA II existieren und damit das Alignment verlässlicher machen. Die erhaltenen Isokonturkarten erlaubten eine Interpretation der Modelle im Hinblick auf die Bedeutung physikochemischer Eigenschaften für die Affinität/Selektivität. Der Vergleich zu qualitativen proteinbasierten Isokonturkarten unterstreicht den komplementären Charakter beider Methoden: Während die ligandbasierten QSAR-Verfahren

implizit teilweise die Struktur der Bindetasche widerspiegeln, aber auch von den Eigenheiten des Trainingsdatensatzes abhängen, vermögen die proteinbasierten Analysen auch Informationen über Bereiche der Bindetasche zu geben, die keine Interaktionen mit Liganden des Trainingsdatensatzes ausbilden.

Ein weiteres Ziel bestand darin, QSAR-Methoden für das Screening größerer Datenbanken zu verwenden. Dies erlaubt die Identifizierung besonders interessanter (d.h. affiner/selektiver) Kandidaten zur Synthese im Sinne einer Leitstrukturoptimierung. Für 3D-QSAR-Methoden musste zunächst ein Protokoll zur Automatisierung des Alignments entwickelt und validiert werden. Es zeigte sich hierbei, dass ein ligandbasiertes Alignment vergleichbare Ergebnisse zu manuellen strukturbasierten Alignmentmethoden erzielt. Die 3D-Modelle erwiesen sich als überlegen im Vergleich zu fragmentbasierten 2D-Methoden oder insbesondere zu den eigenschaftsbasierten 1D-Methoden. Als praktisches Anwendungsbeispiel der entwickelten Modelle wurde eine mehrere tausend Einträge umfassende virtuelle Ligandbibliothek aufgebaut und mit den leistungsfähigsten Modellen bewertet.

Der zweite Teil der Arbeit beinhaltet ein virtuelles Screening nach neuartigen Inhibitoren der Peptiddeformylasen (PDFs) und gehört somit in den Bereich der Leitstrukturfindung. PDFs sind (meist) eisenhaltige Enzyme, die die Deformylierung von in Mitochondrien, Plastiden oder Bakterien synthetisierten Proteinen katalysieren. Ausgehend von Kristallstrukturen potenter PDF-Inhibitoren wurden 3D-Pharmakophormodelle entwickelt und validiert. Diese waren in der Lage, strukturell diverse, aus der Literatur bekannte Inhibitoren zu identifizieren (hinreichende Sensitivität) und gleichzeitig die zu durchsuchenden Datenbanken stark zu reduzieren (hinreichende Spezifität).

Die Pharmakophormodelle wurden zum Screening von Datenbanken kommerziell erhältlicher Moleküle mit wirkstoffartigen Eigenschaften benutzt. Durch Docking und Scoring wurden schließlich aus etwa zwei Millionen Verbindungen elf Substanzen identifiziert und erworben, die einer biologischen Testung unterzogen werden sollen. Erste vorliegende Messergebnisse zeigen, dass mindestens zwei der Substanzen mit einem IC_{50} -Wert von 60 nM bzw. 190 nM potente Inhibitoren der PDF1B aus *E. coli* sind. Dies belegt die Güte der Modelle und des angewendeten Screening-Protokolls. Inhibitoren der PDF könnten Anwendung als Herbizide, Antibiotika und Antimalaria-Therapeutika finden.

9.2 Summary

The modern drug development process is usually divided into the phase of lead discovery and lead optimization. This thesis provides contributions to both aspects.

The first part of this work deals with the establishment and validation of computer models for prediction of affinity and selectivity, thus it supports the area of lead optimization. Selectivity aspects are a major concern in drug development since they influence the risk of undesired side effects. QSAR approaches have been used for prediction of affinity and selectivity parameters using carbonic anhydrases (CAs) as model system. CAs are zinc containing hydrolases which catalyze the reversible hydration of carbon dioxide to bicarbonate and a proton. They are thus involved into various (patho)physiological processes and constitute interesting drug targets. The binding pockets of the numerous isozymes are rather similar in terms of physicochemical properties making the development of selective inhibitors quite difficult.

The investigations were especially focused on 3D-QSAR models. Statistically highly significant and robust models could be established in order to predict affinity and selectivity of sulfonamide inhibitors towards CA I, II, and IV. The investigations revealed that the minor differences in the alignments in the three binding pockets produce only small differences in the statistical outcome. The best models were obtained using the CA II pocket as basis of the alignment, irrespective of the variable to be modeled. Probably the large number of X-ray structures available for CA II is responsible for this effect making the alignment more reliable. The isocontour maps obtained enable an interpretation of the models in terms of physicochemical properties important for affinity/selectivity. The comparison to qualitative protein-based contour maps underlines the complementary character of both approaches: The ligand-based QSAR methods on the one hand implicitly reflect the structure of the binding site partially, on the other hand their outcome depends on the peculiarities of the training set. Protein-based approaches, however, also provide information about regions of the binding pocket which are not occupied by any of the training set ligands.

Another objective was the utilization of QSAR models for database screening. This would allow one to identify the most interesting (i.e. potent/selective) compounds of a database as candidate for synthesis in the context of lead optimization. First, a protocol for automation of ligand alignment had to be developed and validated for

3D-QSAR. The ligand-based alignment procedure applied turned out to yield results comparable to the manual protein-based alignment. The 3D models performed superior to 2D fragment-based or 1D property-based models. As a practical example of application a virtual compound library composed by several thousand entries was designed and ranked using the most predictive models.

The second part of the thesis covered a virtual screening for novel inhibitors of peptide deformylases (PDFs) hence being a task of lead discovery. PDFs are (mostly) ferrous ion containing enzymes cleaving off the formyl group from proteins synthesized in mitochondria, plastids, and eubacteria. 3D-Pharmacophore models have been established and validated using X-ray structures of potent PDF-inhibitors. These pharmacophores were able to identify structurally diverse inhibitors known from literature (sufficient sensitivity) and to strongly reduce the database to be screened (sufficient specificity).

The pharmacophore models were used for screening databases of commercially available druglike compounds. By additional application of docking and scoring eleven out of approximately two million compounds have been identified and purchased. These molecules will be subjected to biological testing. First testing results show that at least two of the selected compounds are potent inhibitors against PDF1B from *E. coli* exhibiting IC_{50} values of 60 nM and 190 nM, respectively. This can be considered as a proof of the quality and suitability of the employed models and the screening protocol. PDF inhibitors could be used as herbicides, antibiotics, and antimalarial therapeutics.

Literaturverzeichnis

- F. Abbate, A. Casini, A. Scozzafava, & C. T. Supuran. Carbonic anhydrase inhibitors: X-ray crystallographic structure of the adduct of human isozyme II with the perfluorobenzoyl analogue of methazolamide. Implications for the drug design of fluorinated inhibitors. *J Enzyme Inhib Med Chem*, 18(4):303–308, Aug 2003.
- F. Abbate, A. Coetzee, A. Casini, S. Ciattini, A. Scozzafava, & C. T. Supuran. Carbonic anhydrase inhibitors: X-ray crystallographic structure of the adduct of human isozyme II with the antipsychotic drug sulpiride. *Bioorg Med Chem Lett*, 14(2):337–341, Jan 2004.
- F. Abbate, C. T. Supuran, A. Scozzafava, P. Orioli, M. T. Stubbs, & G. Klebe. Nonaromatic sulfonamide group as an ideal anchor for potent human carbonic anhydrase inhibitors: role of hydrogen-bonding networks in ligand binding and drug design. *J Med Chem*, 45(17):3583–7., 2002.
- A. Agarwal, P. P. Pearson, E. W. Taylor, H. B. Li, T. Dahlgren, M. Herslof, Y. Yang, G. Lambert, D. L. Nelson, J. W. Regan, & A. R. Martin. Three-dimensional quantitative structure-activity relationships of 5-HT receptor binding data for tetrahydropyridinylindole derivatives: a comparison of the hansch and comfa methods. *J Med Chem*, 36(25):4006–14, 1993.
- Y. Akisawa, I. Nishimori, K. Taniuchi, N. Okamoto, T. Takeuchi, H. Sonobe, Y. Ohtsuki, & S. Onishi. Expression of carbonic anhydrase-related protein CA-RP VIII in non-small cell lung cancer. *Virchows Arch*, 442(1):66–70, Jan 2003.
- B. E. Alber & J. G. Ferry. A carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *Proc Natl Acad Sci U S A*, 91(15):6909–13., 1994.
- AutoQSAR. AutoQSAR. This SVL script is freely available for MOE licensees from the SVL exchange site (<http://svl.chemcomp.com/>).
- J. Baird, T. T., A. Waheed, T. Okuyama, W. S. Sly, & C. A. Fierke. Catalysis and inhibition of human carbonic anhydrase IV. *Biochemistry*, 36(9):2669–78., 1997.
- M. Barboiu, C. T. Supuran, L. Menabuoni, A. Scozzafava, F. Mincione, F. Briganti, & G. Mincione. Carbonic anhydrase inhibitors. Synthesis of topically effective intraocular pressure lowering agents derived from 5-(omega-aminoalkylcarboxamido)-1,3,4-thiadiazole-2-sulfonamide. *J Enzyme Inhib*, 15(1):23–46, 2000.
- G. Barnea, O. Silvennoinen, B. Shaanan, A. M. Honegger, P. D. Canoll, P. D'Eustachio, B. Morse, J. B. Levy, S. Laforgia, & K. Huebner. Identification of a carbonic anhydrase-like domain in the extracellular region of RPTP γ defines a new subfamily of receptor tyrosine phosphatases. *Mol Cell Biol*, 13(3):1497–1506, Mar 1993.
- A. Becker, I. Schlichting, W. Kabsch, D. Groche, S. Schultz, & A. F. Wagner. Iron center, substrate recognition and mechanism of peptide deformylase. *Nat Struct Biol*, 5(12):1053–1058, Dec 1998.

- P. Benedetti, R. Mannhold, G. Cruciani, & G. Ottaviani. GRIND/ALMOND investigations on CysLT1 receptor antagonists of the quinoliny(bridged)aryl type. *Bioorg Med Chem*, 12(13):3607–3617, Jul 2004.
- H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, & P. E. Bourne. The Protein Data Bank. *Nucleic Acids Res*, 28(1):235–42., 2000.
- H. J. Böhm. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des*, 8(3):243–56, 1994.
- M. Böhm. Vergleichende Feldanalysen zur Abschätzung der Bindungsaffinität von Serinprotease-Inhibitoren. Master’s thesis, Universität Regensburg, 1998.
- M. Böhm. *Entwicklung neuer Deskriptoren für vergleichende molekulare Feldanalysen und Verfahren zur Beschreibung der Selektivität von Serinproteasen*. PhD thesis, Philipps-Universität Marburg, 2002.
- M. Böhm, J. Stürzebecher, & G. Klebe. Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J Med Chem*, 42(3):458–77., 1999.
- D. R. Boer, J. Kroon, J. C. Cole, B. Smith, & M. L. Verdonk. SuperStar: comparison of CSD and PDB-based interaction fields as a basis for the prediction of protein-ligand interactions. *J Mol Biol*, 312(1):275–87., 2001.
- P. A. Boriack-Sjodin, R. W. Heck, P. J. Laipis, D. N. Silverman, & D. W. Christianson. Structure determination of murine mitochondrial carbonic anhydrase V at 2.45 Å resolution: implications for catalytic proton transfer and inhibitor design. *Proc Natl Acad Sci U S A*, 92(24):10949–53., 1995.
- J. Borrás, A. Scozzafava, L. Menabuoni, F. Mincione, F. Briganti, G. Mincione, & C. T. Supuran. Carbonic anhydrase inhibitors: synthesis of water-soluble, topically effective intraocular pressure lowering aromatic/heterocyclic sulfonamides containing 8-quinoline-sulfonyl moieties: is the tail more important than the ring? *Bioorg Med Chem*, 7(11):2397–406., 1999.
- A. Boularot, C. Giglione, I. Artaud, & T. Meinel. Structure-activity relationship analysis and therapeutic potential of peptide deformylase inhibitors. *Curr Opin Investig Drugs*, 5(8):809–822, Aug 2004.
- F. Briganti, S. Mangani, P. Orioli, A. Scozzafava, G. Vernaglione, & C. T. Supuran. Carbonic anhydrase activators: X-ray crystallographic and spectroscopic investigations for the interaction of isozymes I and II with histamine. *Biochemistry*, 36(34):10384–92., 1997.
- F. Briganti, S. Mangani, A. Scozzafava, G. Vernaglione, & C. T. Supuran. Carbonic anhydrase catalyzes cyanamide hydration to urea: is it mimicking the physiological reaction? *J Biol Inorg Chem*, 4(5):528–36. 8.htm, 1999.

- F. Briganti, R. Pieratelli, A. Scozzafava, & C. Supuran. Carbonic anhydrase inhibitors. Part 37. Novel classes of isozyme I and II inhibitors and their mechanism of action. Kinetic and spectroscopic investigations on native and cobalt-substituted enzymes. *Eur J Med Chem*, 31:1001–1010, 1996.
- B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, & M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4:187–217, 1983.
- N. Brown & R. A. Lewis. Exploiting QSAR methods in lead optimization. *Curr Opin Drug Discov Devel*, 9(4):419–24, 2006.
- B. L. Bush & J. Nachbar, R. B. Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *J Comput Aided Mol Des*, 7(5):587–619, 1993.
- E. Carosati, R. Mannhold, P. Wahl, J. B. Hansen, T. Fremming, I. Zamora, G. Cianchetta, & M. Baroni. Virtual screening for novel openers of pancreatic K_{ATP} channels. *J Med Chem*, 50(9):2117–2126, May 2007.
- E. Carosati, S. Sciabola, & G. Cruciani. Hydrogen bonding interactions of covalently bonded fluorine atoms: from crystallographic data to a new angular function in the GRID force field. *J Med Chem*, 47(21):5114–5125, Oct 2004.
- A. Casini, A. Scozzafava, F. Mincione, L. Menabuoni, M. A. Ilies, & C. T. Supuran. Carbonic anhydrase inhibitors: water-soluble 4-sulfamoylphenylthioureas as topical intraocular pressure-lowering agents with long-lasting effects. *J Med Chem*, 43(25):4884–92., 2000.
- A. Casini, A. Scozzafava, F. Mincione, L. Menabuoni, & C. T. Supuran. Carbonic anhydrase inhibitors: synthesis of water soluble sulfonamides incorporating a 4-sulfamoylphenylmethylthiourea scaffold, with potent intraocular pressure lowering properties. *J Enzyme Inhib Med Chem*, 17(5):333–43, 2002.
- CATALYST 4.11. Catalyst. version 4.11, Accelrys Inc.: San Diego, CA; <http://www.accelrys.com>.
- S. Chakravarty & K. K. Kannan. Drug-protein interactions. Refined structures of three sulfonamide drug complexes of human carbonic anhydrase I enzyme. *J Mol Biol*, 243(2):298–309., 1994.
- L. C. Chirica, B. Elleby, & S. Lindskog. Cloning, expression and some properties of α -carbonic anhydrase from *Helicobacter pylori*. *Biochim Biophys Acta*, 1544(1-2):55–63., 2001.
- D. W. Christianson & J. D. Cox. Catalysis by metal-activated hydroxide in zinc and manganese metalloenzymes. *Annu Rev Biochem*, 68:33–57, 1999.
- M. Clark, R. D. Cramer, & N. van Opdenbosch. Validation of the General Purpose Tripos 5.2 Force Field. *J Comp Chem*, 10:982–1012, 1989.
- R. D. Clark. OptiSim: an Extended Dissimilarity Selection Method for Finding Diverse Representative Subsets. *J Chem Inf Comput Sci*, 37:1181–88, 1997.

- R. D. Clark. Balancing Representativeness Against Diversity using Optimizable K-Dissimilarity and Hierarchical Clustering. *J Chem Inf Comput Sci*, 38:1079–1086, 1998.
- H. Claussen, C. Buning, M. Rarey, & T. Lengauer. FlexE: efficient molecular docking considering protein structure variations. *J Mol Biol*, 308(2):377–395, Apr 2001.
- S. Clementi. GOLPE 4.5. Multivariate Infometrics Analyses (MIA), 1999.
- E. H. Cox, G. L. McLendon, F. M. Morel, T. W. Lane, R. C. Prince, I. J. Pickering, & G. N. George. The active site structure of *Thalassiosira weissflogii* carbonic anhydrase 1. *Biochemistry*, 39(40):12128–30., 2000.
- R. D. Cramer. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J Am Chem Soc*, 110:5959–5967, 1988.
- R. D. Cramer, S. A. DePriest, D. E. Patterson, & P. Hecht. The Developing Practice of Comparative Molecular Field Analysis. In H. Kubinyi, editor, *3D QSAR in Drug Design: Theory Methods and Applications*, pages 443–485. ESCOM, Leiden, 1993.
- J. D. Cronk, J. A. Endrizzi, M. R. Cronk, W. O'Neill J, & K. Y. Zhang. Crystal structure of *E. coli* β -carbonic anhydrase, an enzyme with an unusual pH-dependent activity. *Protein Sci*, 10(5):911–22., 2001.
- G. Cruciani, editor. *Molecular Interaction Fields*, volume 27 of *Methods and Principles in Medicinal Chemistry*. Wiley-VCH, 1 edition, 2005.
- W. DeLano. The PyMOL Molecular Graphics System, 2005.
- N. Dessalew & P. V. Bharatam. 3D-QSAR and molecular docking study on bisarylmaleimide series as glycogen synthase kinase 3, cyclin dependent kinase 2 and cyclin dependent kinase 4 inhibitors: An insight into the criteria for selectivity. *Eur J Med Chem*, in print, Jan 2007.
- L. M. Dirk, M. A. Williams, & R. L. Houtz. Eukaryotic peptide deformylases. Nuclear-encoded and chloroplast-targeted enzymes in Arabidopsis. *Plant Physiol*, 127(1):97–107, Sep 2001.
- S. J. Dodgson. Inhibition of mitochondrial carbonic anhydrase and ureagenesis: a discrepancy examined. *J Appl Physiol*, 63(5):2134–2141, Nov 1987.
- S. J. Dodgson & K. Cherian. Mitochondrial carbonic anhydrase is involved in rat renal glucose synthesis. *Am J Physiol*, 257(6 Pt 1):E791–E796, Dec 1989.
- J.-M. Dogné, A. Thiry, D. Pratico, B. Masereel, & C. T. Supuran. Dual carbonic anhydrase–cyclooxygenase-2 inhibitors. *Curr Top Med Chem*, 7(9):885–891, 2007.
- A. M. Doweyko. 3D-QSAR illusions. *J Comput Aided Mol Des*, 18(7-9):587–96, 2004.
- D. M. Duda, C. Tu, S. Z. Fisher, H. An, C. Yoshioka, L. Govindasamy, P. J. Laipis, M. Agbandje-McKenna, D. N. Silverman, & R. McKenna. Human carbonic anhydrase III: structural and kinetic study of catalysis and proton transfer. *Biochemistry*, 44(30):10046–10053, Aug 2005.

- D. M. Duda, C. Yoshioka, L. Govindasamy, H. An, C. Tu, D. N. Silverman, & R. McKenna. Crystallization and preliminary X-ray analysis of human carbonic anhydrase III. *Acta Crystallogr D Biol Crystallogr*, 58(Pt 5):849–52, 2002.
- J. L. Durant, B. A. Leland, D. R. Henry, & J. G. Nourse. Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci*, 42(6):1273–80, 2002.
- J. N. Earnhardt, M. Qian, C. Tu, M. M. Lakkis, N. C. Bergenheim, P. J. Laipis, R. E. Tashian, & D. N. Silverman. The catalytic properties of murine carbonic anhydrase VII. *Biochemistry*, 37(30):10837–45., 1998.
- I. Elder, Z. Fisher, P. Laipis, C. Tu, R. McKenna, & D. Silverman. Structural and kinetic analysis of proton shuttle residues in the active site of human carbonic anhydrase III. *Proteins*, 68:337–43, 2007.
- M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, & R. P. Mee. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des*, 11(5):425–445, Sep 1997.
- B. Elleby, B. Sjoblom, C. Tu, D. N. Silverman, & S. Lindskog. Enhancement of catalytic efficiency by the combination of site-specific mutations in a carbonic anhydrase-related protein. *Eur J Biochem*, 267(19):5908–15., 2000.
- A. E. Eriksson, P. M. Kylsten, T. A. Jones, & A. Liljas. Crystallographic studies of inhibitor binding sites in human carbonic anhydrase II: a pentacoordinated binding of the SCN^- -ion to the zinc at high pH. *Proteins*, 4(4):283–293, 1988.
- A. Evers, G. Hessler, H. Matter, & T. Klabunde. Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols. *J Med Chem*, 48(17):5448–65, 2005.
- S. Fieulaine, C. Juillan-Binard, A. Serero, F. Dardel, C. Giglione, T. Meinnel, & J.-L. Ferrer. The crystal structure of mitochondrial (Type 1A) peptide deformylase provides clear guidelines for the design of inhibitors specific for the bacterial forms. *J Biol Chem*, 280(51):42315–42324, Dec 2005.
- J. Free, S. M. & J. W. Wilson. A Mathematical Contribution to Structure-Activity Studies. *J Med Chem*, 53:395–9, 1964.
- K. Fujikawa-Adachi, I. Nishimori, T. Taguchi, & S. Onishi. Human carbonic anhydrase XIV (CA14): cDNA cloning, mRNA expression, and mapping to chromosome 1. *Genomics*, 61(1):74–81., 1999a.
- K. Fujikawa-Adachi, I. Nishimori, T. Taguchi, & S. Onishi. Human mitochondrial carbonic anhydrase VB. cDNA cloning, mRNA expression, subcellular localization, and mapping to chromosome x. *J Biol Chem*, 274(30):21228–33., 1999b.
- J. Gasteiger & M. Marsili. Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. *Tetrahedron*, 36:3219–3228, 1980.

- J. Gasteiger, C. Rudolph, & J. Sadowski. Automatic Generation of 3D Atomic Coordinates for Organic Molecules. *Tetrahedron Comput Methodol*, 3:537–547, 1990.
- M. Gastreich, M. Lilienthal, H. Briem, & H. Claussen. Ultrafast de novo docking combining pharmacophores and combinatorics. *J Comput Aided Mol Des*, 20(12):717–734, Dec 2006.
- C. Geers & G. Gros. Carbon dioxide transport and carbonic anhydrase in blood and muscle. *Physiol Rev*, 80(2):681–715, 2000.
- P. Geladi & B. R. Kowalski. Partial Least Squares Regression: A Tutorial. *Analyt Chim Acta*, 185: 1–17, 1986.
- D. Gelhaar, G. Verkhivker, P. Rejto, D. Fogel, L. Fogel, & S. Freer. Docking conformationally flexible small molecules into a protein binding site through evolutionary programming. In J. McDonnell, R. Reynolds, & D. Fogel, editors, *Proceedings of the Fourth Annual Conference on Evolutionary Programming*, pages 615–627, 1995a.
- D. Gelhaar, G. Verkhivker, P. Rejto, C. Sherman, D. Fogel, L. Fogel, & S. Freer. Molecular recognition of the inhibitor ag-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chemistry & Biology*, 2:317–324, 1995b.
- P. R. Gerber & K. Müller. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J Comput Aided Mol Des*, 9(3):251–68, 1995.
- C. Giglione, A. Boularot, & T. Meinnel. Protein N-terminal methionine excision. *Cell Mol Life Sci*, 61(12):1455–1474, Jun 2004.
- C. Giglione & T. Meinnel. Organellar peptide deformylases: universality of the N-terminal methionine cleavage mechanism. *Trends Plant Sci*, 6(12):566–572, Dec 2001.
- C. Giglione, M. Pierre, & T. Meinnel. Peptide deformylase as a target for new generation, broad spectrum antimicrobial agents. *Mol Microbiol*, 36(6):1197–1205, Jun 2000a.
- C. Giglione, A. Serero, M. Pierre, B. Boisson, & T. Meinnel. Identification of eukaryotic peptide deformylases reveals universality of N-terminal protein processing mechanisms. *EMBO J*, 19(21): 5916–5929, Nov 2000b.
- H. Gohlke. *Entwicklung einer wissensbasierten Bewertungsfunktion zur Struktur- und Affinitätsvorhersage von Protein-Ligand-Komplexen*. PhD thesis, Philipps-Universität Marburg, 2000.
- H. Gohlke, M. Hendlich, & G. Klebe. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol*, 295(2):337–56., 2000a.
- H. Gohlke, M. Hendlich, & G. Klebe. Predicting binding modes, binding affinities and 'hot spots' for protein-ligand complexes using a knowledge-based scoring function. *Perspectives in Drug Discovery and Design*, 20:115–144, 2000b.
- H. Gohlke & G. Klebe. Statistical potentials and scoring functions applied to protein-ligand binding. *Curr Opin Struct Biol*, 11(2):231–5, 2001.

- H. Gohlke & G. Klebe. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chem Int Ed Engl*, 41(15):2644–76, 2002a.
- H. Gohlke & G. Klebe. DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. *J Med Chem*, 45(19):4153–70, 2002b.
- A. Golbraikh, M. Shen, Z. Xiao, Y. D. Xiao, K. H. Lee, & A. Tropsha. Rational selection of training and test sets for the development of validated QSAR models. *J Comput Aided Mol Des*, 17(2-4): 241–53, 2003.
- A. Golbraikh & A. Tropsha. Beware of q^2 ! *Journal of Molecular Graphics and Modelling*, 20(4): 269–76, 2002.
- P. J. Goodford. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem*, 28(7):849–57., 1985.
- D. S. Goodsell & A. J. Olson. Automated docking of substrates to proteins by simulated annealing. *Proteins*, 8(3):195–202, 1990.
- P. A. Greenidge, B. Carlsson, L. G. Bladh, & M. Gillner. Pharmacophores incorporating numerous excluded volumes defined by X-ray crystallographic structure in three-dimensional database searching: application to the thyroid hormone receptor. *J Med Chem*, 41(14):2503–2512, Jul 1998.
- S. Grüneberg, M. T. Stubbs, & G. Klebe. Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J Med Chem*, 45(17):3588–602., 2002.
- D. Groche, A. Becker, I. Schlichting, W. Kabsch, S. Schultz, & A. F. Wagner. Isolation and crystallization of functionally competent *Escherichia coli* peptide deformylase forms containing either iron or nickel in the active site. *Biochem Biophys Res Commun*, 246(2):342–346, May 1998.
- A. Guerri, F. Briganti, A. Scozzafava, C. T. Supuran, & S. Mangani. Mechanism of cyanamide hydration catalyzed by carbonic anhydrase II suggested by cryogenic X-ray diffraction. *Biochemistry*, 39(40):12391–7., 2000.
- J.-P. Guilloteau, M. Mathieu, C. Giglione, V. Blanc, A. Dupuy, M. Chevrier, P. Gil, A. Famechon, T. Meinel, & V. Mikol. The crystal structures of four peptide deformylases bound to the antibiotic actinonin reveal two distinct types: a platform for the structure-based design of antibacterial agents. *J Mol Biol*, 320(5):951–962, Jul 2002.
- S. Handschuh, M. Wagener, & J. Gasteiger. Superposition of three-dimensional chemical structures allowing for conformational flexibility by a hybrid method. *J Chem Inf Comput Sci*, 38(2):220–232, 1998.
- C. Hansch & T. Fujita. ρ - σ - ϕ Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *Journal of the American Chemical Society*, 86:1616–1626, 1964.

- R. W. Heck, P. A. Boriack-Sjodin, M. Qian, C. Tu, D. W. Christianson, P. J. Laipis, & D. N. Silverman. Structure-based design of an intramolecular proton transfer site in murine carbonic anhydrase V. *Biochemistry*, 35(36):11605–11., 1996.
- A. Hillebrecht. 3D-QSAR-Untersuchungen an Norstatininhibitoren des Plasmeptin II mit den Methoden AFMoC, CoMFA und CoMSIA. Master's thesis, Martin-Luther-Universität Halle-Wittenberg, 2005.
- A. Hillebrecht, C. T. Supuran, & G. Klebe. Integrated approach using protein and ligand information to analyze selectivity- and affinity-determining features of carbonic anhydrase isozymes. *ChemMedChem*, 1(8):839–53, 2006.
- M. Hilvo, C. T. Supuran, & S. Parkkila. Characterization and inhibition of the recently discovered carbonic anhydrase isoforms CA XIII, XIV and XV. *Curr Top Med Chem*, 7(9):893–899, 2007.
- M. Hilvo, M. Tolvanen, A. Clark, B. Shen, G. N. Shah, A. Waheed, P. Halmi, M. Hänninen, J. M. Hämäläinen, M. Vihinen, W. S. Sly, & S. Parkkila. Characterization of CA XV, a new GPI-anchored form of carbonic anhydrase. *Biochem J*, 392(Pt 1):83–92, Nov 2005.
- S. A. Hindle, M. Rarey, C. Buning, & T. Lengauer. Flexible docking under pharmacophore type constraints. *J Comput Aided Mol Des*, 16(2):129–49., 2002.
- A. J. Hopfinger, S. Wang, T. J. S., B. Jin, M. Albuquerque, P. J. Madhav, & C. Duraiswami. Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism. *J. Am. Chem. Soc.*, 119: 10509–10524, 1997.
- M. H. Howard, T. Cenizal, S. Gutteridge, W. S. Hanna, Y. Tao, M. Totrov, V. A. Wittenbach, & Y.-J. Zheng. A novel class of inhibitors of peptide deformylase discovered through high-throughput screening and virtual ligand screening. *J Med Chem*, 47(27):6669–6672, Dec 2004.
- S. Huang, Y. Xue, E. Sauer-Eriksson, L. Chirica, S. Lindskog, & B. H. Jonsson. Crystal structure of carbonic anhydrase from *Neisseria gonorrhoeae* and its complex with the inhibitor acetazolamide. *J Mol Biol*, 283(1):301–10, 1998.
- M. Ilies, C. T. Supuran, A. Scozzafava, A. Casini, F. Mincione, L. Menabuoni, M. T. Caproiu, M. Maganu, & M. D. Banciu. Carbonic anhydrase inhibitors: sulfonamides incorporating furan-, thiophene- and pyrrole-carboxamido groups possess strong topical intraocular pressure lowering properties as aqueous suspensions. *Bioorg Med Chem*, 8(8):2145–55., 2000.
- J. J. Irwin & B. K. Shoichet. ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model*, 45(1):177–182, 2005.
- T. M. Iverson, B. E. Alber, C. Kisker, J. G. Ferry, & D. C. Rees. A closer look at the active site of γ -class carbonic anhydrases: high-resolution crystallographic studies of the carbonic anhydrase from *Methanosarcina thermophila*. *Biochemistry*, 39(31):9222–31., 2000.
- G. Jones, P. Willett, & R. C. Glen. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.*, 245(1):43–53, 1995.

- G. Jones, P. Willett, R. C. Glen, A. R. Leach, & R. Taylor. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*, 267(3):727–748, Apr 1997.
- K. M. Jude, S. K. Wright, C. Tu, D. N. Silverman, R. E. Viola, & D. W. Christianson. Crystal structure of F65A/Y131C-methylimidazole carbonic anhydrase V reveals architectural features of an engineered proton shuttle. *Biochemistry*, 41(8):2485–91., 2002.
- K. K. Kannan, M. Ramanadham, & T. A. Jones. Structure, refinement, and function of carbonic anhydrase isozymes: refinement of human carbonic anhydrase I. *Ann N Y Acad Sci*, 429:49–60, 1984.
- M. A. Kastenholtz, M. Pastor, G. Cruciani, E. E. Haaksma, & T. Fox. GRID/CPCA: a new computational tool to design selective ligands. *J Med Chem*, 43(16):3033–44., 2000.
- S. K. Kearsley & G. M. Smith. An alternative method for the alignment of molecular structures: Maximizing electrostatic and steric overlap. *Tetrahedron Comput Methodol*, 3:615–633, 1990.
- K. Köhler, A. Hillebrecht, J. S. Wischeler, A. Innocenti, A. Heine, C. T. Supuran, & G. Klebe. Saccharin Inhibits Carbonic Anhydrases: Possible Explanation for its Unpleasant Metallic Aftertaste. *Angew Chem Int Ed Engl*, 46(40):7697–7699, Aug 2007.
- M. S. Kimber & E. F. Pai. The active site architecture of *Pisum sativum* β -carbonic anhydrase is a mirror image of that of α -carbonic anhydrases. *Embo J*, 19(7):1407–18., 2000.
- M. Kimoto, M. Kishino, Y. Yura, & Y. Ogawa. A role of salivary carbonic anhydrase VI in dental plaque. *Arch Oral Biol*, 51(2):117–122, Feb 2006.
- G. Klebe. Structural Alignment of Molecules. In H. Kubinyi, editor, *3D QSAR in Drug Design: Theory Methods and Applications*, volume 1, pages 173–199. ESCOM, Leiden, 1993.
- G. Klebe. Comparative Molecular Similarity Indices Analysis: CoMSIA. *Perspectives in Drug Discovery and Design*, 12/13/14:87–104, 1998.
- G. Klebe. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today*, 11 (13-14):580–594, Jul 2006.
- G. Klebe & U. Abraham. On the prediction of binding properties of drug molecules by comparative molecular field analysis. *J Med Chem*, 36(1):70–80., 1993.
- G. Klebe & U. Abraham. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *J Comput Aided Mol Des*, 13(1): 1–10., 1999.
- G. Klebe, U. Abraham, & T. Mietzner. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem*, 37(24): 4130–46., 1994.
- G. Klebe & T. Mietzner. A fast and efficient method to generate biologically relevant conformations. *J Comput Aided Mol Des*, 8(5):583–606, 1994.

- J. Kopp & T. Schwede. The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res*, 32(Database issue):D230–4, 2004.
- H. A. Krebs. Inhibition of carbonic anhydrases by sulfonamides. *Biochem J*, 43:525, 1948.
- H. Kubinyi, editor. *3D QSAR in Drug Design: Theory, Methods and Applications*, volume 1. ESCOM, 1993a.
- H. Kubinyi. *QSAR: Hansch Analysis and Related Approaches*, volume 1 of *Methods and Principles in Medicinal Chemistry*. VCH, Weinheim, 1993b.
- H. Kubinyi. Success Stories of Computer-Aided Design. In S. Ekins, editor, *Computer Applications in Pharmaceutical Research and Development*, Wiley Series in Drug Discovery and Development, pages 377–424. John Wiley & Sons, Inc., 2006.
- A. Kumar, K. T. Nguyen, S. Srivathsan, B. Ornstein, S. Turley, I. Hirsh, D. Pei, & W. G. J. Hol. Crystals of peptide deformylase from *Plasmodium falciparum* reveal critical characteristics of the active site for drug design. *Structure*, 10(3):357–367, Mar 2002.
- V. Kumar & K. K. Kannan. Enzyme-substrate interactions. Structure of human carbonic anhydrase I complexed with bicarbonate. *J Mol Biol*, 241(2):226–32., 1994.
- P. Labute. A widely applicable set of descriptors. *J Mol Graph Model*, 18(4-5):464–77, 2000.
- T. W. Lane & F. M. Morel. A biological function for cadmium in marine diatoms. *Proc Natl Acad Sci U S A*, 97(9):4627–4631, Apr 2000.
- T. W. Lane, M. A. Saito, G. N. George, I. J. Pickering, R. C. Prince, & F. M. M. Morel. Biochemistry: a cadmium enzyme from a marine diatom. *Nature*, 435(7038):42, May 2005.
- J. M. Lehtonen, S. Parkkila, D. Vullo, A. Casini, A. Scozzafava, & C. T. Supuran. Carbonic anhydrase inhibitors. Inhibition of cytosolic isozyme XIII with aromatic and heterocyclic sulfonamides: a novel target for the drug design. *Bioorg Med Chem Lett*, 14(14):3757–3762, Jul 2004.
- C. Lemmen & T. Lengauer. Time-efficient flexible superposition of medium-sized molecules. *J Comput Aided Mol Des*, 11(4):357–368, Jul 1997.
- C. Lemmen & T. Lengauer. Computational methods for the structural alignment of molecules. *J Comput Aided Mol Des*, 14(3):215–32., 2000.
- C. Lemmen, T. Lengauer, & G. Klebe. FLEXS: a method for fast flexible ligand superposition. *J Med Chem*, 41(23):4502–20., 1998.
- J. B. Levy, P. D. Canoll, O. Silvennoinen, G. Barnea, B. Morse, A. M. Honegger, J. T. Huang, L. A. Cannizzaro, S. H. Park, & T. Druck. The cloning of a receptor-type protein tyrosine phosphatase expressed in the central nervous system. *J Biol Chem*, 268(14):10573–10581, May 1993.
- A. Liljas, K. Hakansson, B. H. Jonsson, & Y. Xue. Inhibition and catalysis of carbonic anhydrase. Recent crystallographic analyses. *Eur J Biochem*, 219(1-2):1–10., 1994.

- A. Liljas, K. K. Kannan, P. C. Bergstén, I. Waara, K. Fridborg, B. Strandberg, U. Carlbom, L. Järup, S. Lövgren, & M. Petef. Crystal structure of human carbonic anhydrase C. *Nat New Biol*, 235(57): 131–137, Feb 1972.
- S. Lindskog. Structure and mechanism of carbonic anhydrase. *Pharmacol Ther*, 74(1):1–20, 1997.
- S. Lindskog & J. E. Coleman. The catalytic mechanism of carbonic anhydrase. *Proc Natl Acad Sci U S A*, 70(9):2505–8, 1973.
- C. A. Lipinski, F. Lombardo, B. W. Dominy, & P. J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*, 46(1-3):3–26, Mar 2001.
- M. L. López-Rodríguez, M. J. Morcillo, E. Fernández, M. L. Rosado, L. Pardo, & K. Schaper. Synthesis and structure-activity relationships of a new model of arylpiperazines. Study of the 5-HT_{1A}/ α_1 -adrenergic receptor affinity by classical Hansch analysis, artificial neural networks, and computational simulation of ligand recognition. *J Med Chem*, 44(2):198–207, Jan 2001.
- C. J. Lynch, H. Fox, S. A. Hazen, B. A. Stanley, S. Dodgson, & K. F. Lanoue. Role of hepatic carbonic anhydrase in de novo lipogenesis. *Biochem J*, 310 (Pt 1):197–202, Aug 1995.
- E. Magid. The activity of carbonic anhydrases B and C from human erythrocytes and the inhibition of the enzymes by copper. *Scand J Haematol*, 4(4):257–270, 1967.
- R. J. Mallis, B. W. Poland, T. K. Chatterjee, R. A. Fisher, S. Darmawan, R. B. Honzatko, & J. A. Thomas. Crystal structure of S-glutathiolated carbonic anhydrase III. *FEBS Lett*, 482(3):237–41., 2000.
- T. H. Maren. Carbonic anhydrase: chemistry, physiology, and inhibition. *Physiol Rev*, 47(4):595–781., 1967.
- Y. C. Martin. 3D database searching in drug design. *J Med Chem*, 35(12):2145–2154, Jun 1992.
- B. Masereel, S. Rolin, F. Abbate, A. Scozzafava, & C. T. Supuran. Carbonic anhydrase inhibitors: anticonvulsant sulfonamides incorporating valproyl and other lipophilic moieties. *J Med Chem*, 45 (2):312–20., 2002.
- H. Matter & W. Schwab. Affinity and selectivity of matrix metalloproteinase inhibitors: a chemometrical study from the perspective of ligands and proteins. *J Med Chem*, 42(22):4506–23., 1999.
- B. E. Mattioni & P. C. Jurs. Development of quantitative structure-activity relationship and classification models for a set of carbonic anhydrase inhibitors. *J Chem Inf Comput Sci*, 42(1):94–102., 2002.
- M. McGregor & P. Pallai. Clustering of Large Databases of Compounds: Using the MDL Keys as Structural Descriptors. *J Chem Inf Comput Sci*, 37(3):443–448, 1997.
- T. Meinnel. Peptide deformylase of eukaryotic protists: a target for new antiparasitic agents? *Parasitol Today*, 16(4):165–168, Apr 2000.

- L. Menabuoni, A. Scozzafava, F. Mincione, F. Briganti, G. Mincione, & C. T. Supuran. Carbonic anhydrase inhibitors. Water-soluble, topically effective intraocular pressure lowering agents derived from isonicotinic acid and aromatic/heterocyclic sulfonamides: is the tail more important than the ring? *J Enzyme Inhib*, 14(6):457–74, 1999.
- E. Migliavacca. Applied introduction to multivariate methods used in drug discovery. *Mini Rev Med Chem*, 3(8):831–848, Dec 2003.
- F. Mincione, M. Starnotti, L. Menabuoni, A. Scozzafava, A. Casini, & C. T. Supuran. Carbonic anhydrase inhibitors: 4-sulfamoyl-benzenecarboxamides and 4-chloro-3-sulfamoyl-benzenecarboxamides with strong topical antiglaucoma properties. *Bioorg Med Chem Lett*, 11(13):1787–91., 2001.
- G. Mincione, L. Menabuoni, F. Briganti, F. Mincione, A. Scozzafava, & C. T. Supuran. Carbonic anhydrase inhibitors. Part 79. Synthesis of topically acting sulfonamides incorporating GABA moieties in their molecule, with long-lasting intraocular pressure-lowering properties. *Eur J Pharm Sci*, 9(2):185–99, 1999.
- S. Mitsuhashi, T. Mizushima, E. Yamashita, M. Yamamoto, T. Kumasaka, H. Moriyama, T. Ueki, S. Miyachi, & T. Tsukihara. X-ray structure of β -carbonic anhydrase from the red alga, *Porphyridium purpureum*, reveals a novel catalytic site for CO₂ hydration. *J Biol Chem*, 275(8):5521–6., 2000.
- W. E. Müller. Wirkungsmechanismen der atypischen Neuroleptika: Wie wirken Clozapin & Co.? *Pharmazie in unserer Zeit*, 31(6):537–544, 2002.
- MOE 2005.06. MOE version 2005.06. URL <http://www.chemcomp.com>. Chemical Computing Group Inc, Montreal, Quebec, Canada.
- S. Moro, M. Bacilieri, B. Cacciari, C. Bolcato, C. Cusan, G. Pastorin, K.-N. Klotz, & G. Spalluto. The application of a 3D-QSAR (autoMEP/PLS) approach as an efficient pharmacodynamic-driven filtering method for small-sized virtual library: application to a lead optimization of a human A₃ adenosine receptor antagonist. *Bioorg Med Chem*, 14(14):4923–4932, Jul 2006.
- G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, & A. J. Olson. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *Journal of Computational Chemistry*, 19(14):1639–1662, 1998.
- G. M. Morris, D. S. Goodsell, R. Huey, & A. J. Olson. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J Comput Aided Mol Des*, 10(4):293–304., 1996.
- M. Murcia & A. R. Ortiz. Virtual screening with flexible docking and COMBINE-based models. Application to a series of factor Xa inhibitors. *J Med Chem*, 47(4):805–820, Feb 2004.
- Y. Nagao, J. S. Platero, A. Waheed, & W. S. Sly. Human mitochondrial carbonic anhydrase: cDNA cloning, expression, subcellular localization, and mapping to chromosome 16. *Proc Natl Acad Sci U S A*, 90(16):7623–7, 1993.

- T. Naumann & H. Matter. Structural Classification of Protein Kinases Using 3D Molecular Interaction Field Analysis of Their Ligand Binding Sites: Target Family Landscapes. *J Med Chem*, 45(12): 2366–78., 2002.
- I. Nishimori, T. Minakuchi, S. Onishi, D. Vullo, A. Scozzafava, & C. T. Supuran. Carbonic anhydrase inhibitors. DNA cloning, characterization, and inhibition studies of the human secretory isoform VI, a new target for sulfonamide and sulfamate inhibitors. *J Med Chem*, 50(2):381–388, Jan 2007.
- I. Nishimori, D. Vullo, A. Innocenti, A. Scozzafava, A. Mastrolorenzo, & C. T. Supuran. Carbonic anhydrase inhibitors: inhibition of the transmembrane isozyme XIV with sulfonamides. *Bioorg Med Chem Lett*, 15(17):3828–3833, Sep 2005a.
- I. Nishimori, D. Vullo, A. Innocenti, A. Scozzafava, A. Mastrolorenzo, & C. T. Supuran. Carbonic anhydrase inhibitors. The mitochondrial isozyme VB as a new target for sulfonamide and sulfamate inhibitors. *J Med Chem*, 48(24):7860–7866, Dec 2005b.
- Y.-Y. Niu, L.-M. Yang, K.-M. Deng, J.-H. Yao, L. Zhu, C.-Y. Chen, M. Zhang, J.-E. Zhou, T.-X. Shen, H.-Z. Chen, & Y. Lu. Quantitative structure-selectivity relationship for M2 selectivity between M1 and M2 of piperidiny l piperidine derivatives as muscarinic antagonists. *Bioorg Med Chem Lett*, 17(8):2260–2266, Apr 2007.
- S. Oloff, R. B. Mailman, & A. Tropsha. Application of validated QSAR models of D1 dopaminergic antagonists for database mining. *J Med Chem*, 48(23):7322–32, 2005.
- T. I. Oprea & C. L. Waller. Theoretical and Practical Aspects of Three-Dimensional Quantitative Structure-Activity Relationships. In K. B. Lipkowitz & D. B. Boyd, editors, *Reviews in Computational Chemistry*, volume 11, pages 127–182. Wiley-VCH, John Wiley and Sons, Inc., New York, 1997.
- A. R. Ortiz, M. T. Pisabarro, F. Gago, & R. C. Wade. Prediction of drug binding affinities by comparative binding energy analysis. *J Med Chem*, 38(14):2681–2691, Jul 1995.
- M. Pastor & G. Cruciani. A novel strategy for improving ligand selectivity in receptor-based drug design. *J Med Chem*, 38(23):4637–47., 1995.
- M. Pastor, G. Cruciani, I. McLay, S. Pickett, & S. Clementi. GRIND-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J Med Chem*, 43(17):3233–3243, Aug 2000.
- S. Pastorekova, J. Kopacek, & J. Pastorek. Carbonic anhydrase inhibitors and the management of cancer. *Curr Top Med Chem*, 7(9):865–878, 2007.
- S. Pastorekova, S. Parkkila, A. K. Parkkila, R. Opavsky, V. Zelnik, J. Saarnio, & J. Pastorek. Carbonic anhydrase IX, MN/CA IX: analysis of stomach complementary DNA sequence and expression in human and rat alimentary tracts. *Gastroenterology*, 112(2):398–408, 1997.
- PDB-RCSB. RCSB, Protein Data Bank from the Research Collaboratory for Structural Bioinformatics; <http://www.rcsb.org/>.

- D. Pei. Peptide deformylase: a target for novel antibiotics? *Expert Opin Ther Targets*, 5(1):23–40, Feb 2001.
- E. Perola & P. S. Charifson. Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *J Med Chem*, 47(10):2499–2510, May 2004.
- B. Pirard. Peroxisome proliferator-activated receptors target family landscape: a chemometrical approach to ligand selectivity based on protein binding site analysis. *J Comput Aided Mol Des*, 17(11):785–796, Nov 2003.
- B. Pirard & H. Matter. Matrix metalloproteinase target family landscape: a chemometrical approach to ligand selectivity based on protein binding site analysis. *J Med Chem*, 49(1):51–69, Jan 2006.
- A. Popescu, A. Simion, A. Scozzafava, F. Briganti, & C. T. Supuran. Carbonic anhydrase inhibitors. Schiff bases of some aromatic sulfonamides and their metal complexes: towards more selective inhibitors of carbonic anhydrase isozyme IV. *J Enzyme Inhib*, 14(6):407–23, 1999.
- S. Prasanna, E. Manivannan, & S. C. Chaturvedi. QSAR analysis of some fused pyrazoles as selective cyclooxygenase-2 inhibitors: a Hansch approach. *Arch Pharm (Weinheim)*, 337(8):440–444, Aug 2004.
- S. Radestock, M. Böhm, & H. Gohlke. Improving binding mode predictions by docking into protein-specifically adapted potential fields. *J Med Chem*, 48(17):5466–79, 2005.
- S. Ragusa, P. Mouchet, C. Lazennec, V. Dive, & T. Meinnel. Substrate recognition and selectivity of peptide deformylase. Similarities and differences with metzincins and thermolysin. *J Mol Biol*, 289(5):1445–1457, Jun 1999.
- P. T. Rajagopalan, S. Grimme, & D. Pei. Characterization of cobalt(II)-substituted peptide deformylase: function of the metal ion and the catalytic residue Glu-133. *Biochemistry*, 39(4):779–790, Feb 2000.
- P. T. Rajagopalan & D. Pei. Oxygen-mediated inactivation of peptide deformylase. *J Biol Chem*, 273(35):22305–22310, Aug 1998.
- M. Rarey, B. Kramer, & T. Lengauer. Multiple automatic base selection: protein-ligand docking based on incremental construction without manual intervention. *J Comput Aided Mol Des*, 11(4):369–384, Jul 1997.
- M. Rarey, B. Kramer, T. Lengauer, & G. Klebe. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*, 261(3):470–89., 1996.
- M. Rarey & T. Lengauer. A recursive algorithm for efficient combinatorial library docking. *Persp Drug Dis Des*, 20:43–62, 2000.
- G. Renzi, A. Scozzafava, & C. T. Supuran. Carbonic anhydrase inhibitors: topical sulfonamide antiglaucoma agents incorporating secondary amine moieties. *Bioorg Med Chem Lett*, 10(7):673–6., 2000.

- M. Ridderström, I. Zamora, O. Fjellström, & T. B. Andersson. Analysis of selective regions in the active sites of human cytochromes P450, 2C8, 2C9, 2C18, and 2C19 homology models using GRID/CPCA. *J Med Chem*, 44(24):4072–81., 2001.
- M. A. Robien, K. T. Nguyen, A. Kumar, I. Hirsh, S. Turley, D. Pei, & W. G. J. Hol. An improved crystal form of Plasmodium falciparum peptide deformylase. *Protein Sci*, 13(4):1155–1163, Apr 2004.
- I. Salama, C. Hocke, W. Utz, O. Prante, F. Boeckler, H. Hübner, T. Kuwert, & P. Gmeiner. Structure-selectivity investigations of D2-like receptor ligands by CoMFA and CoMSIA guiding the discovery of D3 selective PET radioligands. *J Med Chem*, 50(3):489–500, Feb 2007.
- V. Schnecke & J. Boström. Computational chemistry-driven decision making in lead generation. *Drug Discov Today*, 11(1-2):43–50, Jan 2006.
- G. Schneider, W. Neidhart, T. Giller, & G. Schmid. “Grundgerüstwechsel” (Scaffold-Hopping) durch topologische Pharmakophorsuche: ein Beitrag zum virtuellen Screening. *Angew. Chem.*, 111:3068–3070, 1999.
- L. R. Scolnick. Novel Binding Mode of Hydroxamate Inhibitors to Human Carbonic Anhydrase II. *J Am Chem Soc*, 119:850–851, 1997.
- A. Scozzafava, M. D. Banciu, A. Popescu, & C. T. Supuran. Carbonic anhydrase inhibitors. Inhibition of isozymes I, II and IV by sulfamide and sulfamic acid derivatives. *Journal of Enzyme Inhibition*, 15:443–453, 2000a.
- A. Scozzafava, M. D. Banciu, A. Popescu, & C. T. Supuran. Carbonic anhydrase inhibitors. Synthesis of Schiff bases of hydroxybenzaldehydes with aromatic sulfonamides, and their reactions with arylsulfonyl isocyanates. *Journal of Enzyme Inhibition*, 15:533–546, 2000b.
- A. Scozzafava, L. Menabuoni, F. Mincione, F. Briganti, G. Mincione, & C. T. Supuran. Carbonic anhydrase inhibitors. Synthesis of water-soluble, topically effective, intraocular pressure-lowering aromatic/heterocyclic sulfonamides containing cationic or anionic moieties: is the tail more important than the ring? *J Med Chem*, 42(14):2641–50., 1999.
- A. Scozzafava, L. Menabuoni, F. Mincione, F. Briganti, G. Mincione, & C. T. Supuran. Carbonic anhydrase inhibitors: perfluoroalkyl/aryl-substituted derivatives of aromatic/heterocyclic sulfonamides as topical intraocular pressure-lowering agents with prolonged duration of action. *J Med Chem*, 43(23):4542–51., 2000c.
- A. Scozzafava, F. Mincione, L. Menabuoni, & C. T. Supuran. Carbonic anhydrase inhibitors: topically acting antiglaucoma sulfonamides incorporating phthaloyl and phthalimido moieties. *Drug Des Discov*, 17(4):337–48, 2001.
- A. Scozzafava & C. T. Supuran. Carbonic anhydrase inhibitors. Arylsulfonylureido- and arylureido-substituted aromatic and heterocyclic sulfonamides: towards selective inhibitors of carbonic anhydrase isozyme I. *J Enzyme Inhib*, 14(5):343–63, 1999.

- A. Scozzafava & C. T. Supuran. Carbonic anhydrase and matrix metalloproteinase inhibitors: sulfonylated amino acid hydroxamates with MMP inhibitory properties act as efficient inhibitors of CA isozymes I, II, and IV, and N-hydroxysulfonamides inhibit both these zinc enzymes. *J Med Chem*, 43(20):3677–87., 2000a.
- A. Scozzafava & C. T. Supuran. Carbonic Anhydrase Inhibitors: synthesis of N-Morpholythiocarbonylsulfenylamino aromatic/heterocyclic sulfonamides and their interaction with Isozymes I, II and IV. *Bioorganic & Medicinal Chemistry Letters*, 10:1117–1120, 2000b.
- M. Shen, C. Beguin, A. Golbraikh, J. P. Stables, H. Kohn, & A. Tropsha. Application of predictive QSAR models to database mining: identification and experimental validation of novel anticonvulsant compounds. *J Med Chem*, 47(9):2356–64, 2004.
- G. D. Simone & C. T. Supuran. Antiobesity carbonic anhydrase inhibitors. *Curr Top Med Chem*, 7(9):879–884, 2007.
- W. Sippl. Development of biologically active compounds by combining 3D QSAR and structure-based design methods. *J Comput Aided Mol Des*, 16(11):825–830, Nov 2002.
- W. S. Sly & P. Y. Hu. Human carbonic anhydrases and carbonic anhydrase deficiencies. *Annu Rev Biochem*, 64:375–401, 1995.
- G. M. Smith, R. S. Alexander, D. W. Christianson, B. M. McKeever, G. S. Ponticello, J. P. Springer, W. C. Randall, J. J. Baldwin, & C. N. Habecker. Positions of His-64 and a bound water in human carbonic anhydrase II upon binding three structurally related inhibitors. *Protein Sci*, 3(1):118–25., 1994.
- H. K. Smith, R. P. Beckett, J. M. Clements, S. Doel, S. P. East, S. B. Launchbury, L. M. Pratt, Z. M. Spavold, W. Thomas, R. S. Todd, & M. Whittaker. Structure-activity relationships of the peptide deformylase inhibitor BB-3497: modification of the metal binding group. *Bioorg Med Chem Lett*, 12(24):3595–3599, Dec 2002.
- K. S. Smith & J. G. Ferry. A plant-type (β -class) carbonic anhydrase in the thermophilic methanarchaeon *Methanobacterium thermoautotrophicum*. *J Bacteriol*, 181(20):6247–6253, Oct 1999.
- K. S. Smith & J. G. Ferry. Prokaryotic carbonic anhydrases. *FEMS Microbiol Rev*, 24(4):335–66., 2000.
- C. A. Sotriffer & I. Dramburg. “In situ cross-docking” to simultaneously address multiple targets. *J Med Chem*, 48(9):3122–3125, May 2005.
- C. A. Sotriffer, H. Gohlke, & G. Klebe. Docking into knowledge-based potential fields: a comparative evaluation of DrugScore. *J Med Chem*, 45(10):1967–70., 2002.
- S. S. Spicer, Z. H. Ge, R. E. Tashian, D. J. Hazen-Martin, & B. A. Schulte. Comparative distribution of carbonic anhydrase isozymes III and II in rodent tissues. *Am J Anat*, 187(1):55–64, 1990.
- M. Stahl & M. Rarey. Detailed analysis of scoring functions for virtual screening. *J Med Chem*, 44(7):1035–42, 2001.

- J. S. Strobl, M. Cassell, S. M. Mitchell, C. M. Reilly, & D. S. Lindsay. Scriptaid and suberoylanilide hydroxamic acid are histone deacetylase inhibitors with potent anti-*Toxoplasma gondii* activity in vitro. *J Parasitol*, 93(3):694–700, Jun 2007.
- P. Strop, K. S. Smith, T. M. Iverson, J. G. Ferry, & D. C. Rees. Crystal structure of the cab-type β -class carbonic anhydrase from the archaeon *Methanobacterium thermoautotrophicum*. *J Biol Chem*, 276(13):10299–305., 2001.
- G. H. Su, T. A. Sohn, B. Ryu, & S. E. Kern. A novel histone deacetylase inhibitor identified by high-throughput transcriptional screening of a compound library. *Cancer Res*, 60(12):3137–3142, Jun 2000.
- C. Supuran, A. Scozzafava, & J. Conway, editors. *Carbonic Anhydrase - Its Inhibitors and Activators*. CRC Enzyme Inhibitors Series. CRC Press, 2004.
- C. T. Supuran. Carbonic anhydrase inhibitors - Part 53. Synthesis of substituted-pyridinium derivatives of aromatic sulfonamides: The first non-polymeric membrane-impermeable inhibitors with selectivity for isozyme IV. *Eur J Med Chem*, 33:577–594, 1998.
- C. T. Supuran. Carbonic anhydrase inhibitors - Part 57: Quantum chemical QSAR of a group of 1,3,4-thiadiazole- and 1,3,4-thiadiazoline disulfonamides with carbonic anhydrase inhibitory properties. *Eur J Med Chem*, 34:41–50, 1999.
- C. T. Supuran, F. Briganti, L. Menabuoni, G. Mincione, F. Mincione, & A. Scozzafava. Carbonic anhydrase inhibitors - part 78. Synthesis of water-soluble sulfonamides incorporating β -alanyl moieties, possessing long lasting-intraocular pressure lowering properties via the topical route. *Eur J Med Chem*, 35(3):309–21., 2000a.
- C. T. Supuran, C. W. Conroy, & T. H. Maren. Is cyanate a carbonic anhydrase substrate? *Proteins*, 27(2):272–278, Feb 1997.
- C. T. Supuran & A. Scozzafava. Carbonic anhydrase inhibitors-Part 94. 1,3,4-thiadiazole-2- sulfonamidederivatives as antitumor agents? *Eur J Med Chem*, 35(9):867–74., 2000a.
- C. T. Supuran & A. Scozzafava. Carbonic anhydrase inhibitors and their therapeutic potential. *Exp. Opin. Ther. Patents*, 10(5):575–600, 2000b.
- C. T. Supuran & A. Scozzafava. Carbonic anhydrase inhibitors: aromatic sulfonamides and disulfonamides act as efficient tumor growth inhibitors. *J Enzyme Inhib*, 15(6):597–610, 2000c.
- C. T. Supuran & A. Scozzafava. Carbonic Anhydrase Inhibitors. *Curr Med Chem: Imm, Endoc and Metab Agents*, 1:61–97, 2001.
- C. T. Supuran & A. Scozzafava. Carbonic anhydrases as targets for medicinal chemistry. *Bioorg Med Chem*, 15(13):4336–4350, Jul 2007.
- C. T. Supuran, A. Scozzafava, & A. Casini. Carbonic anhydrase inhibitors. *Med Res Rev*, 23(2): 146–89., 2003.

- C. T. Supuran, A. Scozzafava, M. A. Ilies, & F. Briganti. Carbonic anhydrase inhibitors: synthesis of sulfonamides incorporating 2,4,6-trisubstituted-pyridinium-ethylcarboxamido moieties possessing membrane-impermeability and in vivo selectivity for the membrane-bound (CA IV) versus the cytosolic (CA I and CA II) isozymes. *J Enzyme Inhib*, 15(4):381–401, 2000b.
- C. T. Supuran, A. Scozzafava, B. C. Jurca, & M. A. Ilies. Carbonic anhydrase inhibitors. Part 49. Synthesis of substituted ureido- and thioureido derivatives of aromatic/heterocyclic sulfonamides with increased affinities for isozyme I. *European Journal of Medicinal Chemistry*, 33:83–93, 1998.
- C. T. Supuran, A. Scozzafava, L. Menabuoni, F. Mincione, F. Briganti, & G. Mincione. Carbonic anhydrase inhibitors. Part 71. Synthesis and ocular pharmacology of a new class of water-soluble, topically effective intraocular pressure lowering sulfonamides incorporating picolinoyl moieties. *Eur J Pharm Sci*, 8(4):317–28, 1999.
- E. Svastová, N. Zilka, M. Zat'ovicová, A. Gibadulinová, F. Ciampor, J. Pastorek, & S. Pastoreková. Carbonic anhydrase IX reduces E-cadherin-mediated adhesion of MDCK cells via interaction with β -catenin. *Exp Cell Res*, 290(2):332–345, Nov 2003.
- SYBYL 7.0. SYBYL molecular modeling package, 2004. URL <http://www.tripos.com>. SYBYL® 7.0 Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
- SYBYL 7.1. SYBYL molecular modeling package, 2005. URL <http://www.tripos.com>. SYBYL® 7.1 Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
- R. E. Tashian, D. Hewett-Emmett, N. D. Carter, & N. C. H. Bergenhenn. Carbonic anhydrase (CA)-related proteins (CA-RPs), and transmembrane proteins with CA or CA-RP domains. In W. R. Chegwidden, N. D. Carter, & Y. H. Edwards, editors, *The Carbonic Anhydrases - New Horizons*, pages 105–120. Birkhäuser Verlag, Basel, 2000.
- C. Temperini, A. Innocenti, A. Mastrolorenzo, A. Scozzafava, & C. T. Supuran. Carbonic anhydrase inhibitors. Interaction of the antiepileptic drug sulthiame with twelve mammalian isoforms: kinetic and X-ray crystallographic studies. *Bioorg Med Chem Lett*, 17(17):4866–4872, Sep 2007.
- A. J. Tervo, T. H. Nyrönen, T. Rönkkö, & A. Poso. Comparing the quality and predictiveness between 3D QSAR models obtained from manual and automated alignment. *J Chem Inf Comput Sci*, 44(3):807–816, 2004.
- U. Thibaut, G. Folkers, G. Klebe, H. Kubinyi, A. Merz, & D. Rognan. Recommendations for CoMFA Studies and 3D QSAR Publications. In H. Kubinyi, editor, *3D QSAR in Drug Design: Theory Methods and Applications*, volume 1, pages 711–716. ESCOM, Leiden, 1993.
- A. Thiry, J.-M. Dogné, C. T. Supuran, & B. Masereel. Carbonic anhydrase inhibitors as anticonvulsant agents. *Curr Top Med Chem*, 7(9):855–864, 2007.
- N. Triballeau, F. Acher, I. Brabet, J. P. Pin, & H. O. Bertrand. Virtual screening workflow development guided by the receiver operating characteristic curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Med Chem*, 48(7):2534–47, 2005.

- N. Triballeau, H. O. Bertrand, & F. Acher. Are You Sure You Have a Good Model? In T. Langer & R. D. Hoffmann, editors, *Pharmacophores and Pharmacophore Searches*, volume 32 of *Methods and Principles in Medicinal Chemistry*, pages 325–364. WILEY-VCH, Weinheim, 2006.
- B. C. Tripp, K. Smith, & J. G. Ferry. Carbonic anhydrase: new insights for an ancient enzyme. *J Biol Chem*, 276(52):48615–8., 2001.
- A. Tropsha. Application of Predictive QSAR Models to Database Mining. In T. I. Oprea, editor, *Cheminformatics in Drug Discovery*, volume 23 of *Methods and Principles in Medicinal Chemistry*, pages 437–455. WILEY-VCH, Weinheim, 2004.
- C. Tu, M. Qian, J. N. Earnhardt, P. J. Laipis, & D. N. Silverman. Properties of intramolecular proton transfer in carbonic anhydrase III. *Biophys J*, 74(6):3182–9., 1998.
- C. K. Tu, D. N. Silverman, C. Forsman, B. H. Jonsson, & S. Lindskog. Role of histidine 64 in the catalytic mechanism of human carbonic anhydrase II studied with a site-specific mutant. *Biochemistry*, 28(19):7913–8, 1989.
- J. I. Vandenberg, B. D. Walker, & T. J. Campbell. HERG K⁺ channels: friend and foe. *Trends Pharmacol Sci*, 22(5):240–246, May 2001.
- A. Vedani & M. Dobler. 5D-QSAR: the key for simulating induced fit? *J Med Chem*, 45(11):2139–2149, May 2002.
- A. Vedani, M. Dobler, & M. A. Lill. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *J Med Chem*, 48(11):3700–3703, Jun 2005.
- H. F. G. Velec, H. Gohlke, & G. Klebe. DrugScore^{CSD}-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J Med Chem*, 48(20):6296–6303, Oct 2005.
- M. L. Verdonk, V. Berdini, M. J. Hartshorn, W. T. M. Mooij, C. W. Murray, R. D. Taylor, & P. Watson. Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J Chem Inf Comput Sci*, 44(3):793–806, 2004.
- M. L. Verdonk, G. Chessari, J. C. Cole, M. J. Hartshorn, C. W. Murray, J. W. M. Nissink, R. D. Taylor, & R. Taylor. Modeling water molecules in protein-ligand docking using GOLD. *J Med Chem*, 48(20):6504–6515, Oct 2005.
- M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, & R. D. Taylor. Improved protein-ligand docking using GOLD. *Proteins*, 52(4):609–623, Sep 2003.
- M. L. Verdonk, J. C. Cole, & R. Taylor. SuperStar: a knowledge-based approach for identifying interaction sites in proteins. *J Mol Biol*, 289(4):1093–108., 1999.
- M. L. Verdonk, J. C. Cole, P. Watson, V. Gillet, & P. Willett. SuperStar: improved knowledge-based interaction fields for protein binding sites. *J Mol Biol*, 307(3):841–59., 2001.

- R. M. Vitale, C. Pedone, P. Amodeo, J. Antel, M. Wurl, A. Scozzafava, C. T. Supuran, & G. D. Simone. Molecular modeling study for the binding of zonisamide and topiramate to the human mitochondrial carbonic anhydrase isoform VA. *Bioorg Med Chem*, 15(12):4152–4158, Jun 2007.
- D. Vullo, J. Voipio, A. Innocenti, C. Rivera, H. Ranki, A. Scozzafava, K. Kaila, & C. T. Supuran. Carbonic anhydrase inhibitors. Inhibition of the human cytosolic isozyme VII with aromatic and heterocyclic sulfonamides. *Bioorg Med Chem Lett*, 15(4):971–976, Feb 2005.
- R. C. Wade, K. J. Clark, & P. J. Goodford. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J Med Chem*, 36(1):140–7., 1993.
- R. C. Wade & P. J. Goodford. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J Med Chem*, 36(1):148–56., 1993.
- A. Weber. *Kristallstrukturanalyse und Entwicklung von Computermodellen zur Beschreibung der Selektivität von Enzymen am Beispiel der Carboanhydrase*. Phd, Philipps-Universität, 2004.
- A. Weber, M. Böhm, C. T. Supuran, A. Scozzafava, C. A. Sotriffer, & G. Klebe. 3D QSAR selectivity analyses of carbonic anhydrase inhibitors: insights for the design of isozyme selective inhibitors. *J Chem Inf Model*, 46(6):2737–2760, 2006.
- D. A. Whittington, J. H. Grubb, A. Waheed, G. N. Shah, W. S. Sly, & D. W. Christianson. Expression, assay, and structure of the extracellular domain of murine carbonic anhydrase XIV: implications for selective inhibition of membrane-associated isozymes. *J Biol Chem*, 279(8):7223–7228, Feb 2004.
- D. A. Whittington, A. Waheed, B. Ulmasov, G. N. Shah, J. H. Grubb, W. S. Sly, & D. W. Christianson. Crystal structure of the dimeric extracellular domain of human carbonic anhydrase XII, a bitopic membrane protein overexpressed in certain cancer tumor cells. *Proc Natl Acad Sci U S A*, 98(17):9545–50., 2001.
- B. L. Wilkinson, L. F. Bornaghi, T. A. Houston, A. Innocenti, D. Vullo, C. T. Supuran, & S.-A. Poulsen. Inhibition of membrane-associated carbonic anhydrase isozymes IX, XII and XIV with a library of glycoconjugate benzenesulfonamides. *Bioorg Med Chem Lett*, 17(4):987–992, Feb 2007.
- I. H. Witten & E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann series in data management systems. Morgan Kaufmann Publishers, San Fransisco, 2nd edition, 2005.
- S. Wold. Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, 2:37–52, 1987.
- S. Wold, E. Johansson, & M. Cocchi. PLS - Partial Least-Squares Projections to Latent Structures. In H. Kubinyi, editor, *3D QSAR in Drug Design: Theory Methods and Applications*, pages 523–550. ESCOM, Leiden, 1993.

- G. Wong, K. F. Koehler, P. Skolnick, Z. Q. Gu, S. Ananthan, P. Schonholzer, W. Hunkeler, W. Zhang, & J. M. Cook. Synthetic and computer-assisted analysis of the structural requirements for selective, high-affinity ligand binding to diazepam-insensitive benzodiazepine receptors. *J Med Chem*, 36(13): 1820–30, 1993.
- Z. Yuan, J. Trias, & R. White. Deformylase as a novel antibacterial target. *Drug Discov Today*, 6 (18):954–961, Sep 2001.
- M. Zentgraf, J. Fokkens, & C. A. Sotriffer. Addressing protein flexibility and ligand selectivity by “in situ cross-docking”. *ChemMedChem*, 1(12):1355–1359, Dec 2006.
- Q. Y. Zhang, J. Wan, X. Xu, G. F. Yang, Y. L. Ren, J. J. Liu, H. Wang, & Y. Guo. Structure-based rational quest for potential novel inhibitors of human HMG-CoA reductase by combining CoMFA 3D QSAR modeling and virtual screening. *J Comb Chem*, 9(1):131–138, 2007a.
- T. Zhang, J.-H. Zhou, L.-W. Shi, R.-X. Zhu, & M.-B. Chen. 3D-QSAR studies with the aid of molecular docking for a series of non-steroidal FXR agonists. *Bioorg Med Chem Lett*, 17(8):2156–2160, Apr 2007b.
- W. Zheng & A. Tropsha. Novel variable selection quantitative structure–property relationship approach based on the k-nearest-neighbor principle. *J Chem Inf Comput Sci*, 40(1):185–94, 2000.
- J. Zupan & J. Gasteiger. Quantitative Structure-Activity Relationships. In *Neural Networks in Chemistry and Drug Design*, chapter 13, pages 219–243. Wiley-VCH, 2 edition, 1999.
- J. Závada, Z. Zavadová, J. Pastorek, Z. Biesová, J. Jezek, & J. Velek. Human tumour-associated cell adhesion protein MN/CA IX: identification of M75 epitope and of the region mediating cell adhesion. *Br J Cancer*, 82(11):1808–1813, Jun 2000.

Danksagung

Mein herzlicher Dank gilt:

- Herrn Prof. Dr. GERHARD KLEBE für seine Unterstützung, für die interessanten Themenstellungen der Projekte, und die Freiheit, die Richtung der Projekte selbst beeinflussen zu dürfen. Ganz besonders möchte ich mich für seine ermutigenden Worte im Hinblick auf Vortragssituationen bedanken.
- Herrn Prof. Dr. CHRISTOPH A. SOTRIFTER für die hilfreichen Diskussionen und Hinweise bei verschiedenen Detailfragen im Bereich der Computerchemie und die Übernahme der Funktion des Zweitgutachters.
- Herrn Prof. Dr. MARTIN SCHLITZER für die Übernahme der Funktion des Vorsitzenden der Prüfungskommission sowie die fruchtbare Zusammenarbeit bei verschiedenen Projekten und seine Offenheit für Vorschläge aus der Computerchemie.
- Herrn Prof. Dr. KLAUS KUSCHINSKY, Prof. Dr. KLAUS-HEINRICH RÖHM sowie allen Mitgliedern der Prüfungskommission für die Bereitschaft, trotz des weihnachtlichen Termins an der Disputation teilzunehmen.
- GERD NEUDERT und SVEN SIEBLER für die kompetente Administration des Computernetzwerkes der AG Klebe; ferner für ihre und auch Dr. HOLGER STEUBERS Gesellschaft und gemeinsame Mahlzeiten zu nächtlicher bzw. morgendlicher Stunde.
- Dr. CHRISTOF GERLACH (MODELLING-BÄR) für interessante fachliche und nichtfachliche Diskussionen sowie das kritische Lesen der Doktorarbeit.
- Der FUN-GENERATION (PATRICK PFEFFER (PEPPI), TINA RITSCHER (DINO), MARTIN SIPPEL (WOMI), ANDREAS SPITZMÜLLER) für den Spaß und die Beratung in Fragen des Lifestyles.
- ANGELA SCHOLZ für die aufmerksame Hilfe in vielen organisatorischen und bürokratischen Angelegenheiten. Außerdem möchte ich mich für viele erbauliche Worte bedanken.

- BERNHARD BAUM als meinem Ex-Bürogenossen und LISA ENGLERT für die Auflockerung der Büroatmosphäre und interessante Gespräche verschiedenster Art.
- Herrn Prof. Dr. THIERRY LANGER und seiner Gruppe für die Ermöglichung eines Gastaufenthaltes an der UNIVERSITÄT INNSBRUCK, um vertiefte Kenntnisse im Bereich der Pharmakophormodellierung zu erlangen.
- Dr. ANDREAS EVERS, Dr. ALEXANDER WEBER und Dr. PAUL CZODROWSKI für die Überlassung der L^AT_EX-Vorlage.
- Der gesamten AG Klebe für die angenehme und abwechslungsreiche Arbeitsatmosphäre.

Aus der vorliegenden Arbeit sind folgende Posterbeiträge, Vorträge und Publikationen hervorgegangen:

Posterbeiträge:

- Hillebrecht, A., Schlitzer, M., Meinnel, T., Langer, T., Klebe, G., *Virtual Screening for Novel Inhibitors of Peptide Deformylases*. Poster präsentiert auf dem gemeinsamen deutsch-schweizerischen GDCh/DPhG-Meeting „Frontiers in Medicinal Chemistry“, Berlin (2007).
- Hillebrecht, A., Supuran, C.T., Klebe, G., *Elucidation of Affinity- and Selectivity-Determining Features of Carbonic Anhydrase I, II, and IV by Protein- and Ligand-Based Approaches*. Poster präsentiert auf dem jährlichen GDCh-Meeting „Frontiers in Medicinal Chemistry“, Frankfurt (2006).

Vorträge:

- *Modelling Affinity- and Selectivity-Determining Features Using Different 2D- and 3D-QSAR Approaches*. Vortrag präsentiert auf dem Molecular Modelling Workshop, Erlangen (2006).

Aufsätze:

- Hillebrecht, A., Klebe, G., *The Use of 3D QSAR Models for Database Screening: A Feasibility Study*. *J. Chem. Inf. Model.* 2007, angenommen.
- Köhler, K., Hillebrecht, A., Schulze Wischeler, J., Innocenti, A., Heine, A., Supuran, C.T., Klebe, G., *Saccharin Inhibits Carbonic Anhydrases: Possible Explanation for its Unpleasant Metallic Aftertaste*. *Angew. Chem. Int. Ed. Engl.* 2007, 46(40): 7697-7699.
- Azim-Zadeh, O., Hillebrecht, A., Linne, U., Marahiel, M.A., Klebe, G., Lingelbach, K., Nyalwidhe, J., *The Use of Biotin Derivatives to Probe Conformational Changes in Proteins*. *J. Biol. Chem.* 2007, 282(30): 21609-21617.
- Hillebrecht, A., Supuran, C.T., Klebe, G., *Integrated Approach Using Protein and Ligand Information to Analyze Selectivity- and Affinity-Determining Features of Carbonic Anhydrase Isozymes*. *ChemMedChem* 2006, 1, 539-552.

Erklärung

Ich versichere, dass ich meine Dissertation

*Modellierung von Metalloenzymen: 3D-QSAR-Untersuchungen an
Carboanhydrase-Isoenzymen und virtuelles Screening nach
Peptiddeformylase-Inhibitoren*

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch in keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 26. November 2007

(Alexander Hillebrecht)

Lebenslauf

Alexander Hillebrecht

Geburtstag:	27. Oktober 1978
Geburtsort:	Friedberg
Staatsangehörigkeit:	deutsch
1985 – 1987	Grundschule in Friedberg-Bruchenbrücken
1987 – 1989	Grundschule in Niddatal-Assenheim
1989 – 1998	<i>St.-Lioba-Gymnasium</i> Bad Nauheim
Jun. 1998	Abitur
Okt. 1998	Immatrikulation im Studiengang Pharmazie an der <i>J.-W.-Goethe-Universität</i> Frankfurt (Main)
Aug. 2000	1. Staatsexamen
Apr. 2003	2. Staatsexamen
Mai 2003 – Okt. 2003	1. Hälfte des praktischen Jahres; Diplomarbeit am Institut für Pharmazeutische Chemie der <i>Philipps-Universität</i> Marburg unter Anleitung von Prof. Dr. G. Klebe; Titel: „3D-QSAR-Untersuchungen an Norstatin-Inhibitoren des Plasmepsin II mit den Methoden AFMoC, CoMFA und CoMSIA“
Nov. 2003 – Apr. 2004	2. Hälfte des praktischen Jahres in der <i>Birken-Apotheke</i> , Lahntal
Mai 2004	3. Staatsexamen
Jun. 2004	Approbation als Apotheker
Jun. 2004	Beginn der Arbeiten zur vorliegenden Dissertation unter Betreuung von Prof. Dr. G. Klebe am Institut für Pharmazeutische Chemie der <i>Philipps-Universität</i> Marburg
Jun. 2004 – Dez. 2007	Wissenschaftlicher Angestellter am Institut für Pharmazeutische Chemie der <i>Philipps-Universität</i> Marburg; Betreuung des Studentenpraktikums im fünften Semester Pharmazie